The Role of Physics in Urban Science

ETA Seminar

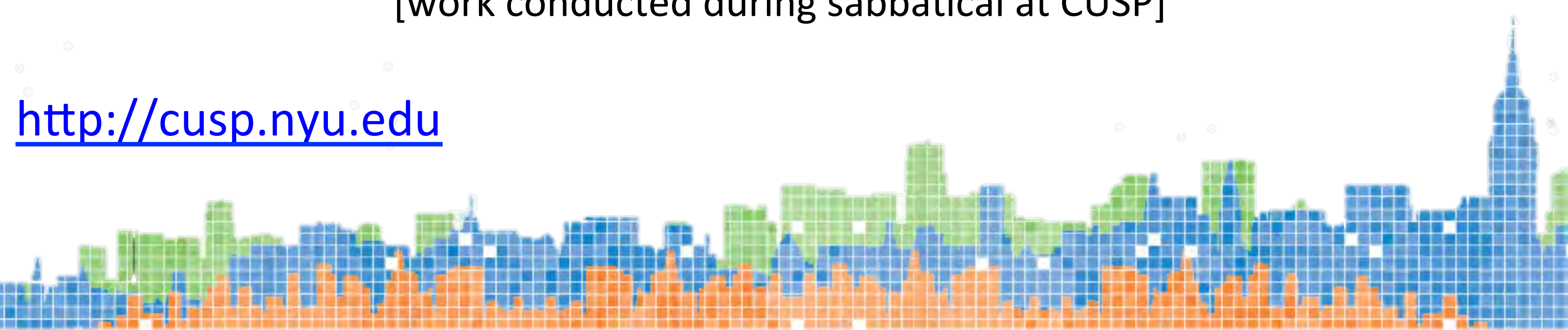
January 29, 2014

Jonathan Wurtele

UC Berkeley and LBNL

[work conducted during sabbatical at CUSP]

http://cusp.nyu.edu



CUSP

The Center for Urban Science and Progress (CUSP) is a unique public-private research center that uses New York City as its laboratory and classroom to help cities around the world become more productive, livable, equitable, and resilient. CUSP observes, analyzes, and models cities to optimize outcomes, prototype new solutions, formalize new tools and processes, and develop new expertise/ experts. These activities will make CUSP the world's leading authority in the emerging field of "Urban Informatics."

Steve Koonin, 2012

The CUSP vision includes New York City as its laboratory: A national lab for cities.

What does it mean to instrument a city?

Infrastructure

Condition, operations

Environment

Meteorology, pollution, noise, flora, fauna

People

Relationships, location, economic /communications activities, health, nutrition, opinions, organizations, …

Properly acquired, integrated, and analyzed data potentially leads to:

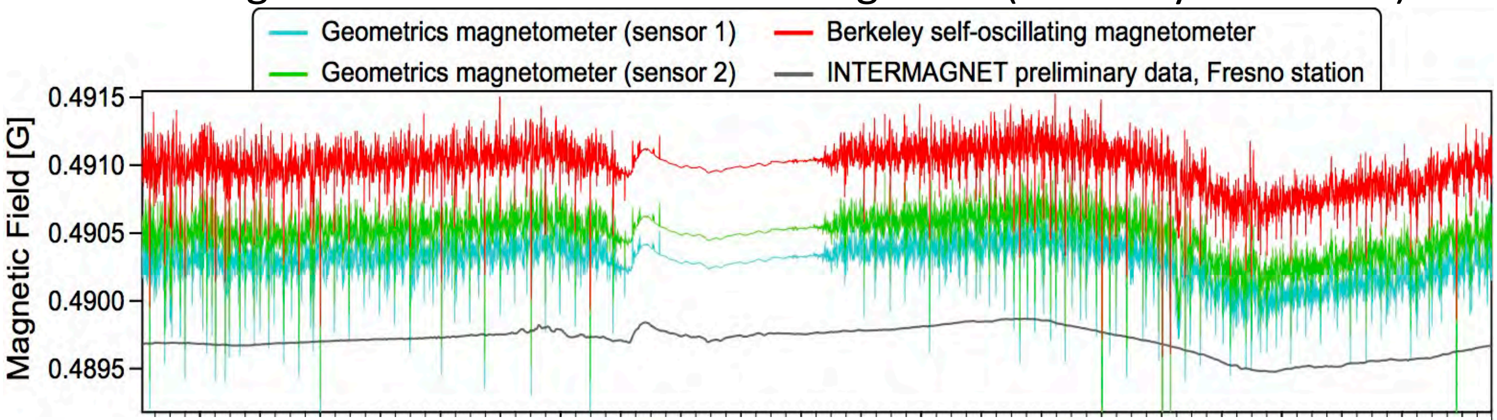
- Better (and more efficient) operations, better planning, better policy: Does a policy work? Can policy goals be analytically described and evaluated? Do we have enough data to understand the impact of a policy?
- New models for citizen engagement
- Enabling the private sector to develop new services for citizens, governments, firms
- Enabling a revolution in the social sciences

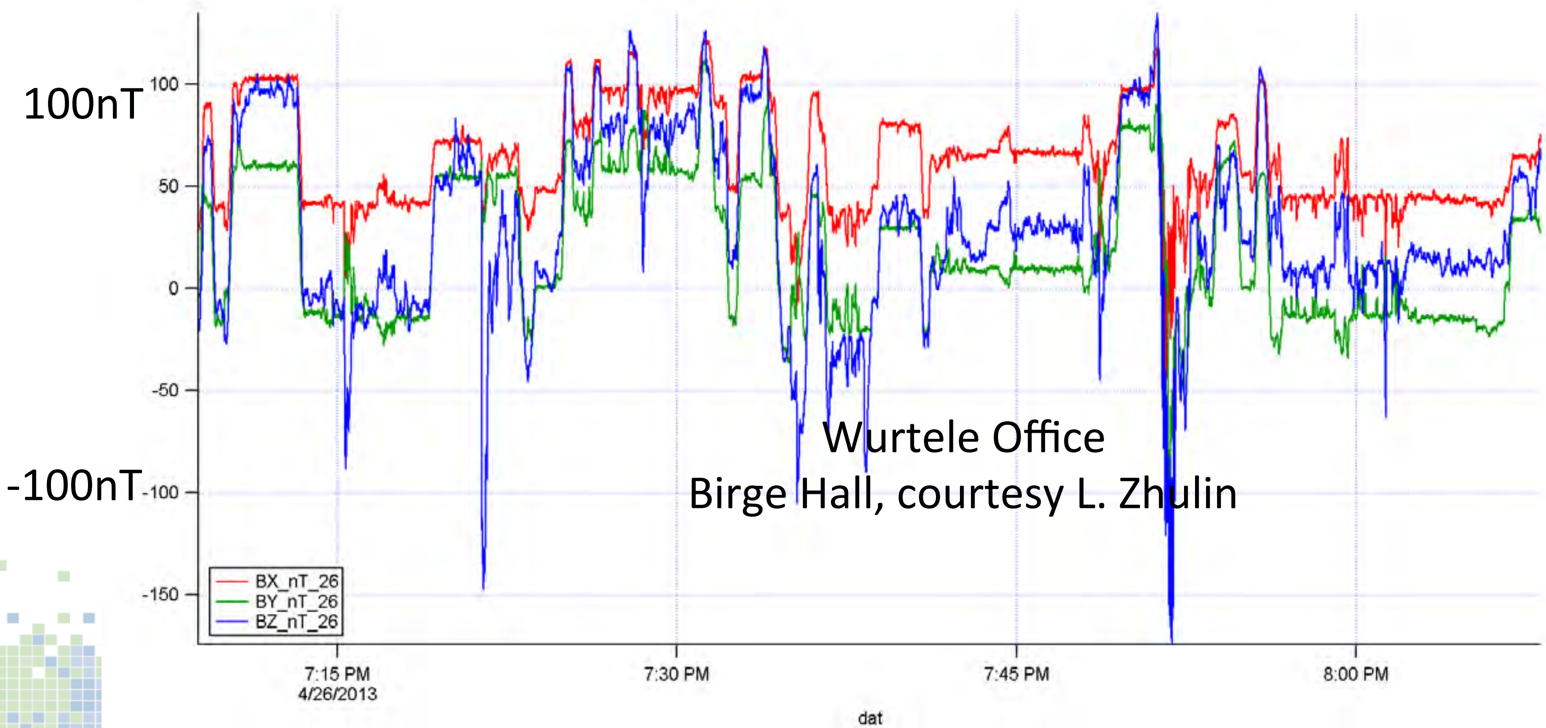
"***Get your facts first, then you can distort them as you please.***"

Mark Twain

Magnet Sensors

Magnetic field at the UC botanical garden (Courtesy D. Budker)





Urban Data Sources

- **City and citizen data flows**
  - Administrative records (census, permits, …)
  - Transactions (sales, communications, …)
  - Operational (traffic, transit, utilities, health system, …)
  - Social media (Twitter feeds, blog posts, Facebook, …)

- **Sensors**
  - Personal (location, activity, physiological)
  - Fixed *in situ* sensors
  - Crowd sourcing (mobile phones, …)
  - Choke points (people, vehicles)

- **Opportunities for "novel" sensor technologies**
  - Visible, infrared and spectral imagery
  - RADAR, LIDAR
  - Gravity and magnetic
  - Seismic, acoustic
  - Ionizing radiation, biological, chemical
  - …

These new data sources offer unprecedented:
granularity
variety
coverage
timeliness

Macro View: Scaling laws for cities

West, Bittencourt, Batty, others

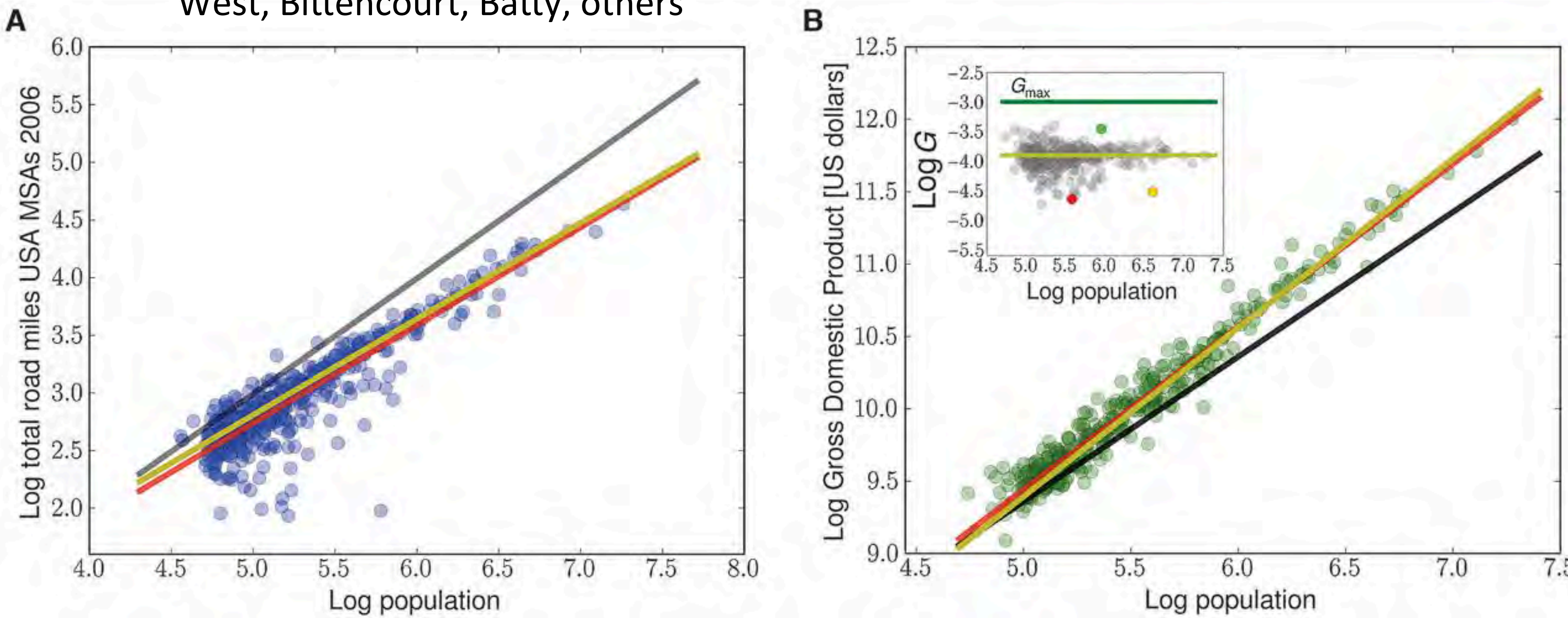


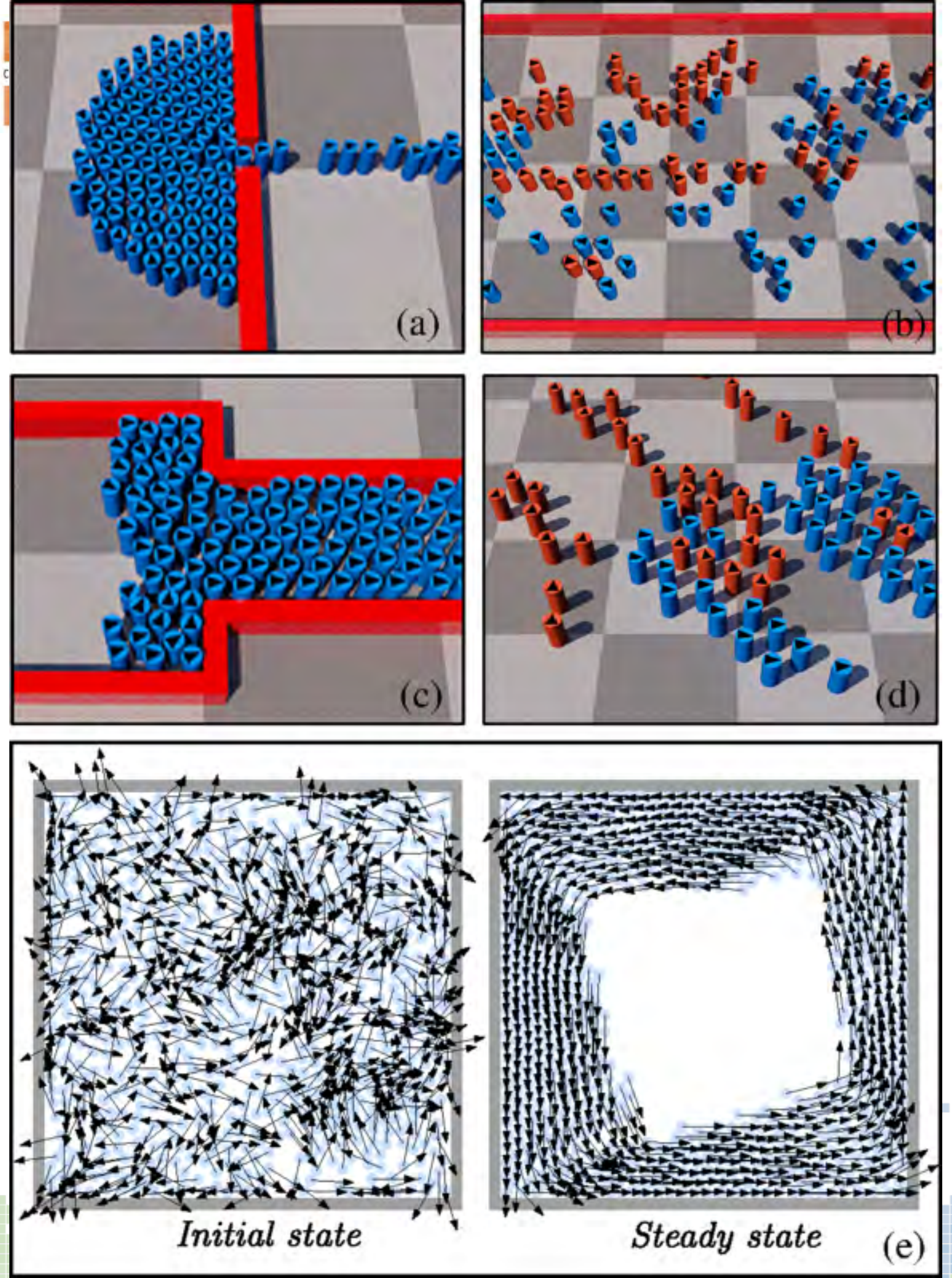
Fig. 1 Scaling of urban infrastructure and socioeconomic output.(A) Total lane miles (volume) of roads in U.S. metropolitan areas (MSAs) in 2006 (blue dots).

L M A Bettencourt Science 2013;340:1438-1441

Science AAAS

Published by AAAS

Micro-View: Scaling laws for small-scale dynamics

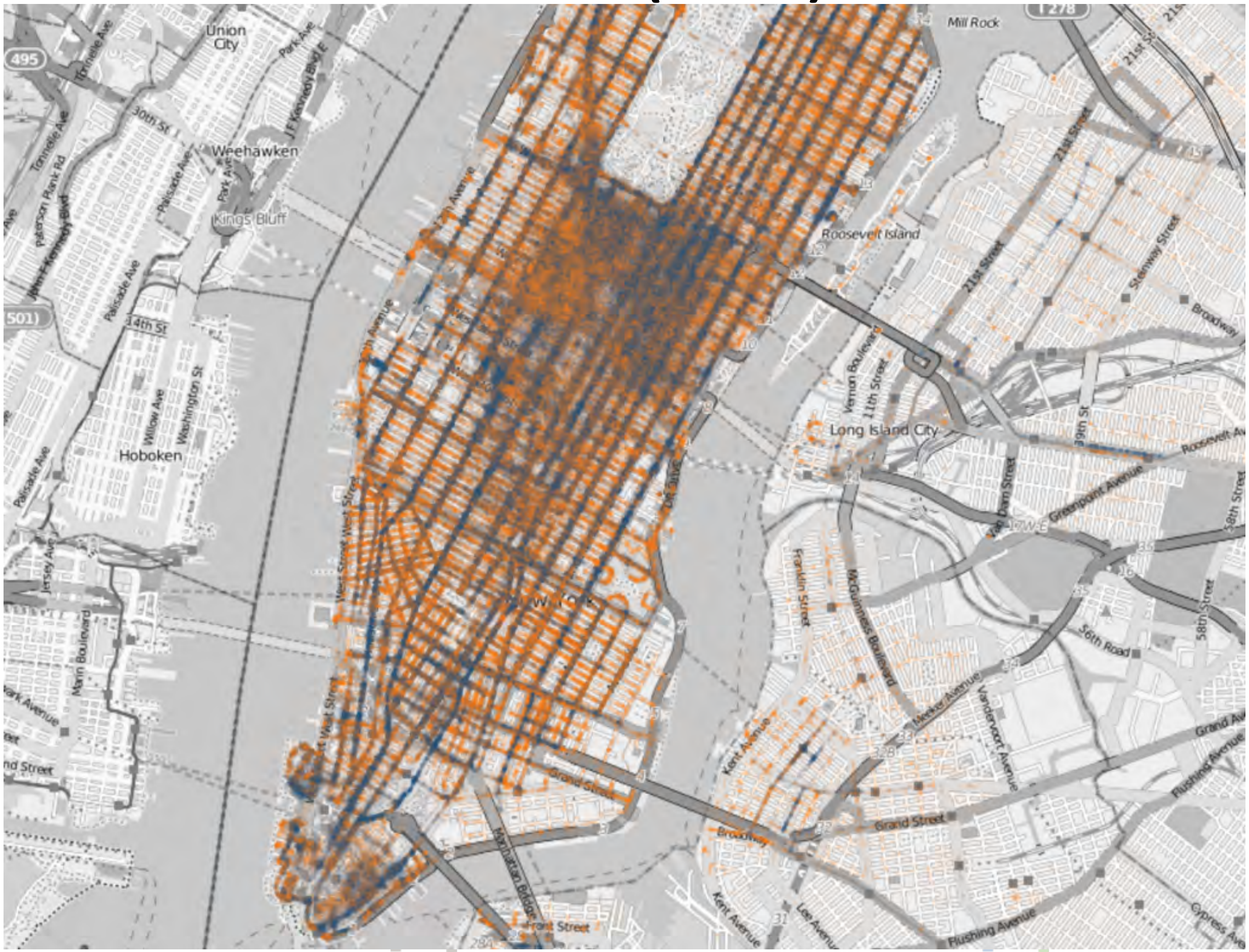


Agent based models of pedestrian dynamics

Karamouzas, Skinner, Guy
PRL 2014

simple power-law interaction
projected time to a potential collision.
Applies to a wide variety of situations, speeds, and densities.

Visualization of Spatiotemporal Events
Taxi (TLC) GPS Data



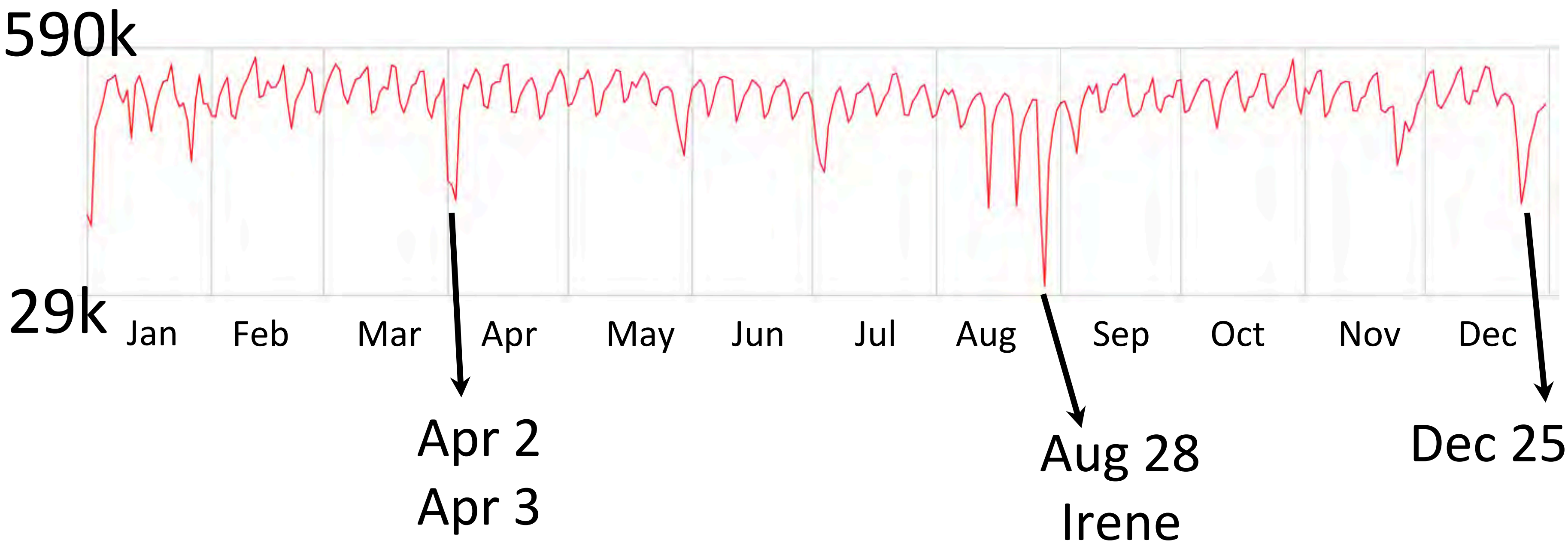
Drop-off

Pick-up

Most drop-off's occur on the avenues, most pick-up's on the streets

Lauro Lins, Fernando Chirigati, Nivan Ferreira, Claudio Silva, and Juliana Freire, NYU- Poly
(Data obtained from TLC on June 6, 2012)

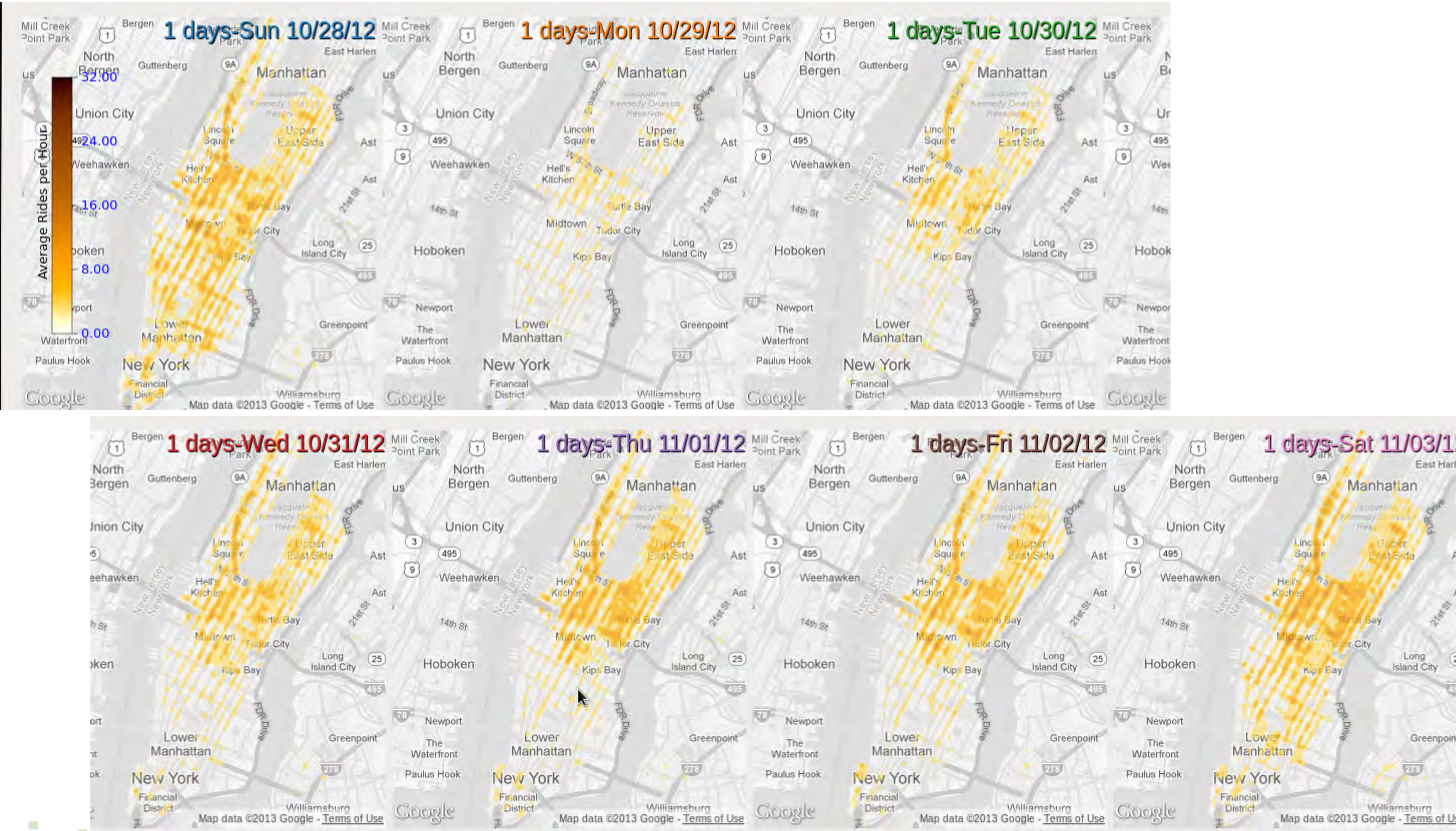
NYC Taxi Rides by Day in 2011



Lauro Lins, Fernando Chirigati, Nivan Ferreira, Claudio Silva, and Juliana Freire, NYU- Poly
(Data obtained from TLC on June 6, 2012)

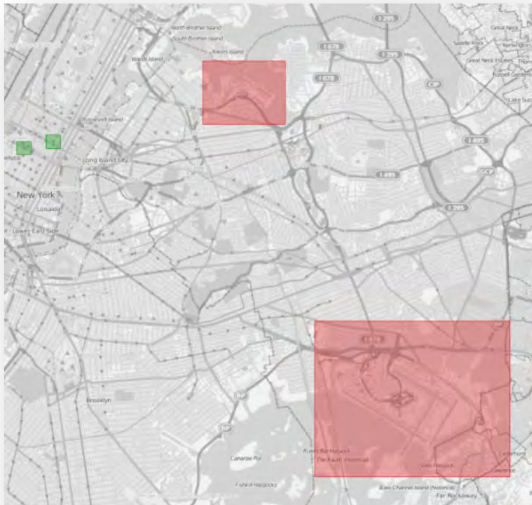
Taxi Rides in Manhattan, October 28 – November 3, 2012

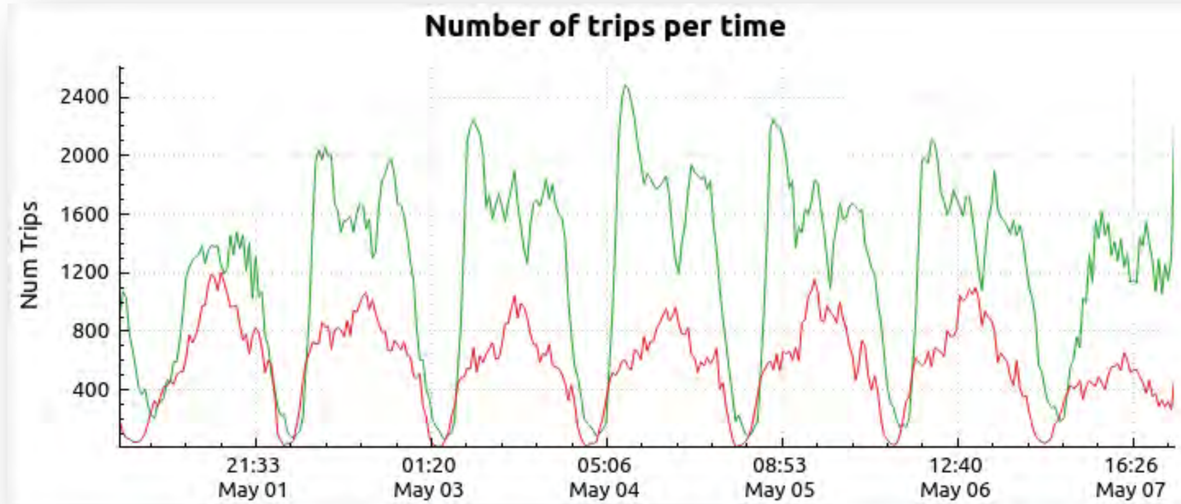
(Superstorm Sandy)



Juliana Freire, Claudio Silva, et al, NYU-Poly

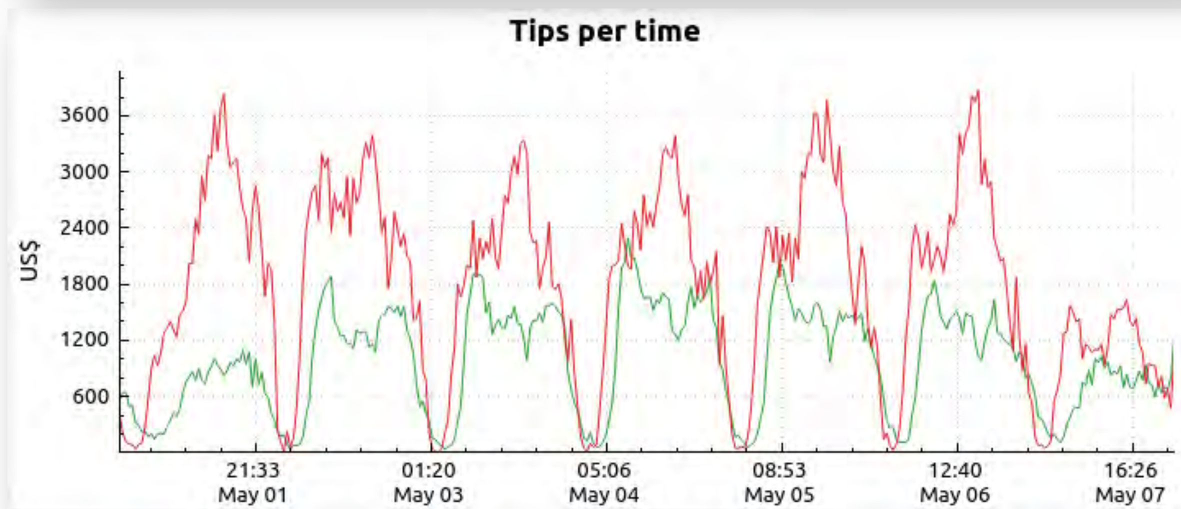
Studying Taxi Patterns





Train Stations

Airports

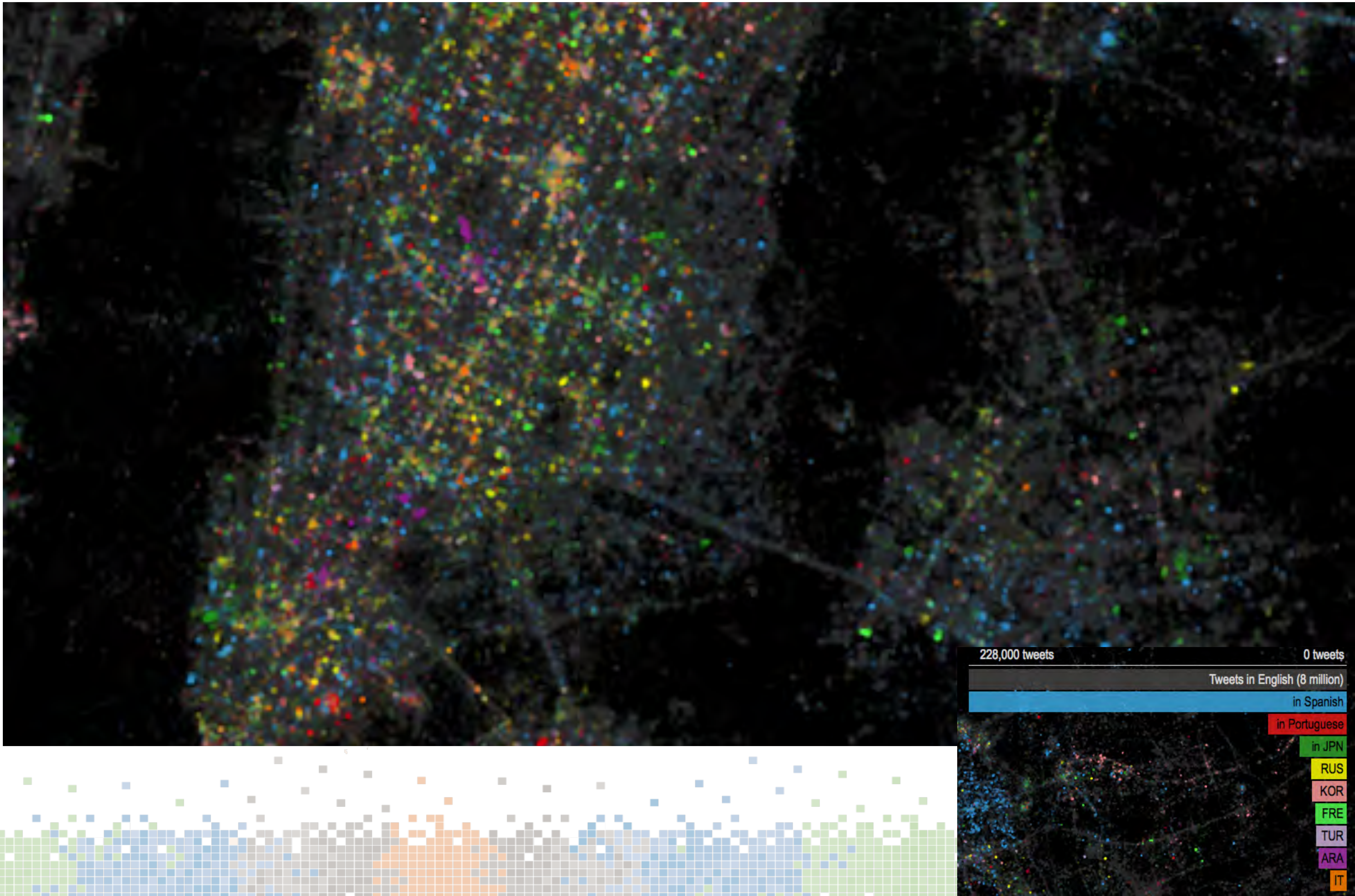


May 1^{st} – 7^{th} 2011

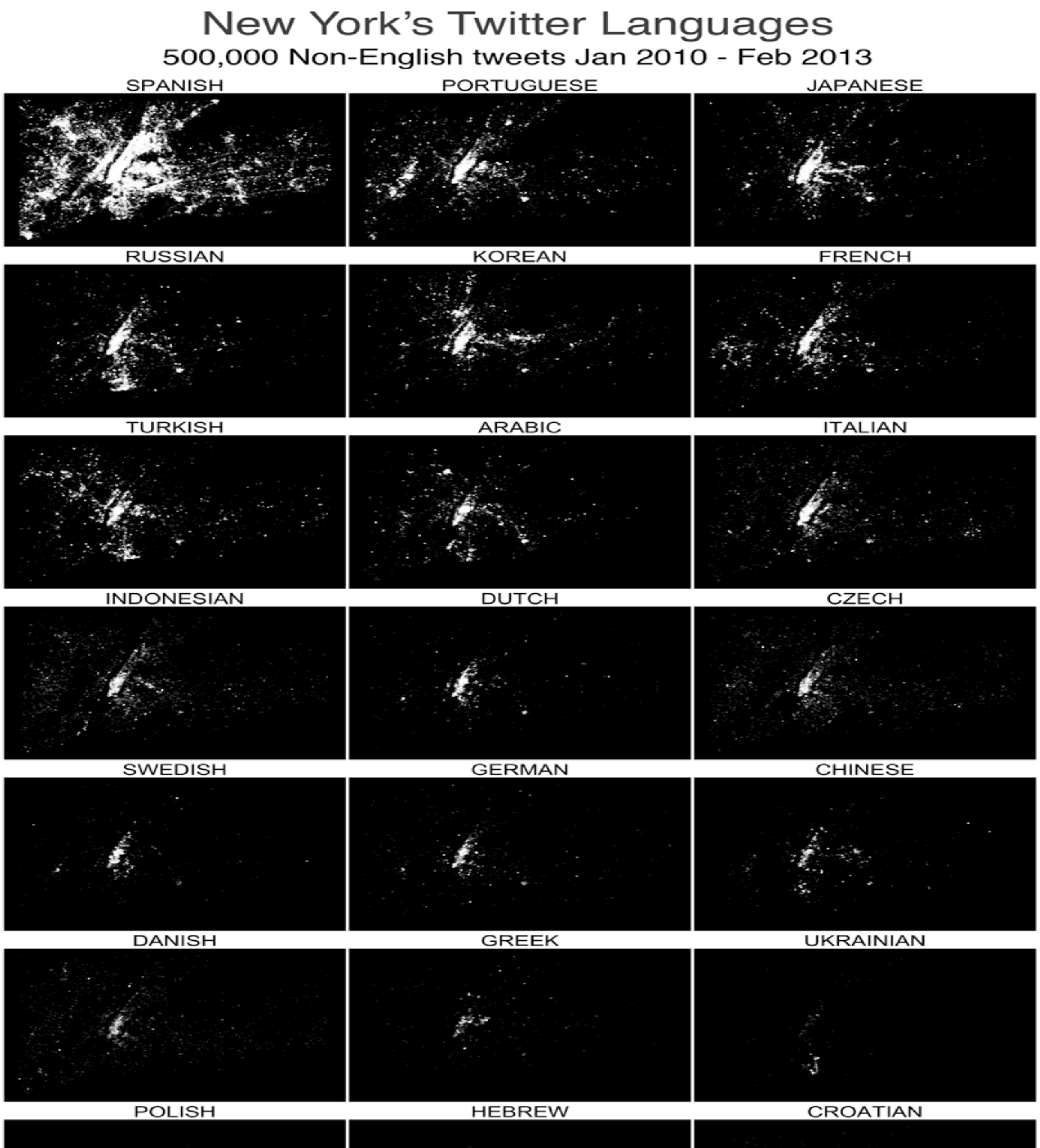
3.6 Million Trips

http://ny.spatial.ly/

Twitter



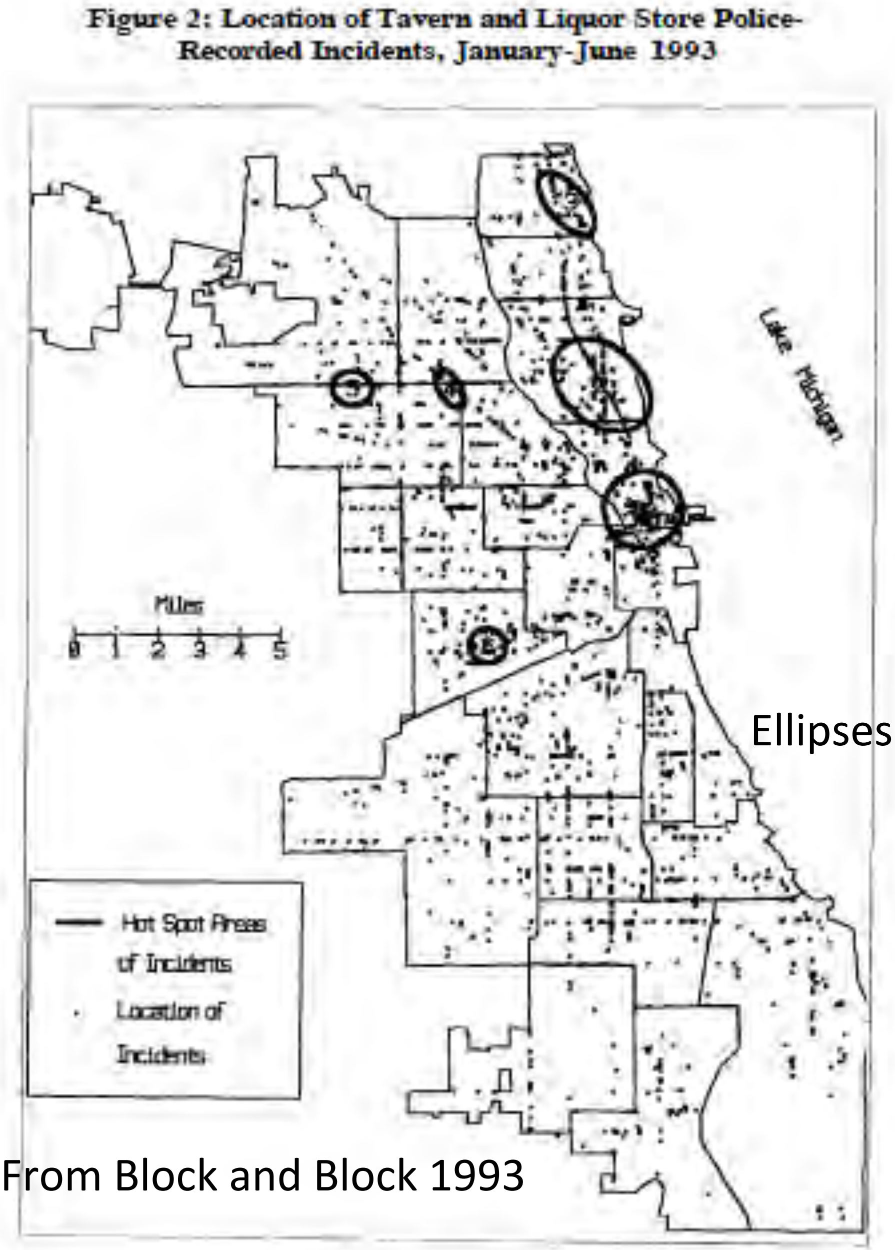
CASASUse Browser



Big data and crime

1. Methods for predicting crimes: forecast places and times with an increased risk of crime.
   – Optimization of resource allocation.
2. Identify individuals at risk as future offenders
   – For preventive measures
3. Assist solving crimes by pattern recognition.
4. Identify individuals or groups at risk of being victims of a crime.
5. Improve recidivism rates.

1993 crime map



From Block and Block 1993

Can now include:

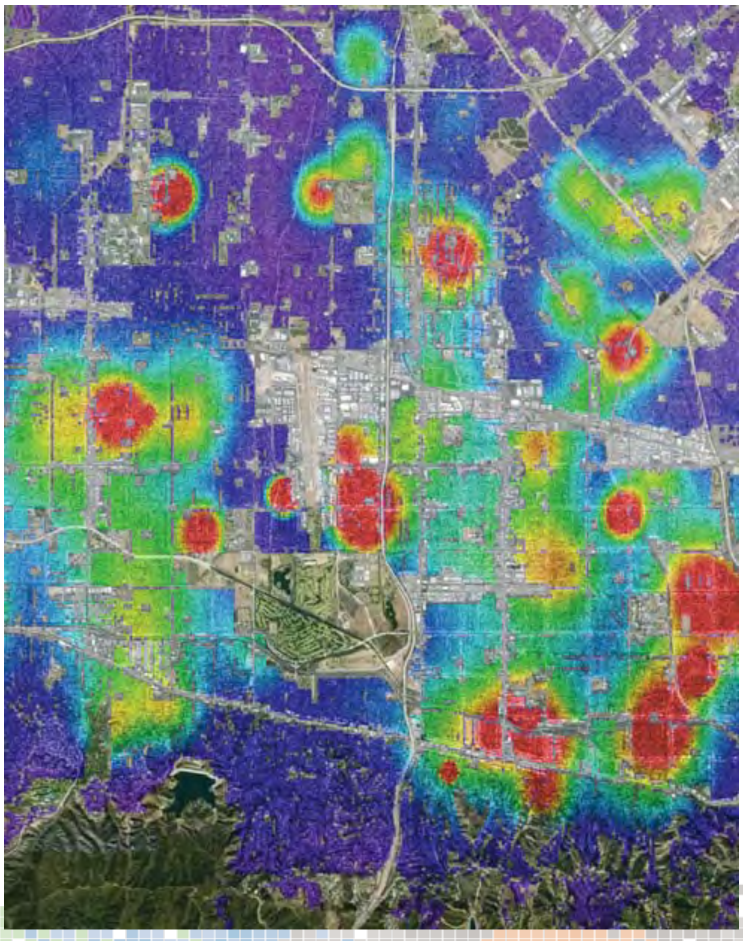
Multiple correlative data sets
Temporal and spatial statistics
Variation with time of day, day of week, temperature, season, income, …
sophisticated algorithms
Much more data
More reliable data (?)

Ellipses indicate regions of higher crime

The math behind the scene of the crime,

Martin B. Short Physics Today 2014

CRIME HEAT MAP



Hotspots from Partial Differential equations.

These hotspots move dynamically and the
PDE is derived from a plausible agent based model.

But no claims that this produces operation advantage.

Later the 'earthquake' model was developed by the same UCLAgroup (**Bertozzi**, Short, Mohler, et al.)

Mathematical methods for predictive policing

Self-Exciting Point Process Modeling of Crime

G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita

Highly clustered event sequences are observed in certain types of crime data, such as burglary and gang violence, due to crime-specific patterns of criminal behavior. Similar clustering patterns are observed by seismologists, as earthquakes are well known to increase the risk of subsequent earthquakes, or aftershocks, near the location of an initial event. Space–time clustering is modeled in seismology by self-exciting point processes and the focus of this article is to show that these methods are well suited for criminological applications. We first review self ... s Angeles Police Departmen ... purpose we use a fully nonp ... oral trends in the backgroun

KEY WOR

Criminological research has shown that crime can spread through local environments via a contagion-like process (Johnson 2008). For example, burglars will repeatedly attack clusters of nearby targets because local vulnerabilities are well known to the offenders (Bernasco and Nieuwbeerta 2005). A gang shooting may incite waves of retaliatory violence in the local set space (territory) of the rival gang (Tita and Ridgeway 2007; Cohen and Tita 1999). The local, contagious spread of crime leads to the formation of crime clusters in space and time.

The infectiousness of crime is like that of the plague.
[Napoleon Bonaparte](#)

Two opinions

Crime is terribly revealing.
Try and vary your methods as you will,
your tastes, your habits, your attitude of mind,
and your soul is revealed by your actions.
Agatha Christie

The world of crime is a last refuge of the authentic, uncorrupted, spontaneous event.
Daniel J. Boorstin

Prediction of crime

- Predicting probabilities not events
- Input:
  - Historical data
  - (possibly) correlative data
    - 311/911
    - Demographics
    - Graffiti
- Output:
  - Actionable map of probability of crime
- Methods:
  - Geospatial mapping
  - K-means or similar cluster analysis
  - Background and anomaly detection
  - Model for tendency of (some) crimes to cluster in time

Very useful overview of the field:
Predictive Policing, RAND Report
http://www.rand.org/pubs/research_reports/RR233.html

Predpol.com

online talks, e.g., http://vimeo.com/50315082

Problem

Identify areas at increased risk

Using historical crime data

Goel, Rao and Shroff
Submitted 2014
https://5harad.com/papers/frisky.pdf

Stop-and-frisks

(a)

Homicides

(b)

Demographics

(c)

How do you measure success?
Urban science is **NOT** physics

- How do know if your model worked?
  - Postdiction vs prediction.
  - Quantitative vs Qualitative
- What is your benchmark?
  - Is what you are doing cost effective?
    - Don't bring a tank if you only need a squirt gun. (Don't resort to fancy models if simplea are just as good).
  - Is your method better than what is already being done?

- Think about tests. They may not be easy to realize
  - **Example**: compare crime change over some length of time in precincts that use one method vs another. Problem: criminality is dynamic. individual enthusiasm can impact results.

CUSP Facilities Being Developed

- **Data warehouse (access policies, technologies, infrastructure)**
- **Urban Observatory**
- **"SimNYC"**
- **Quantified Community**

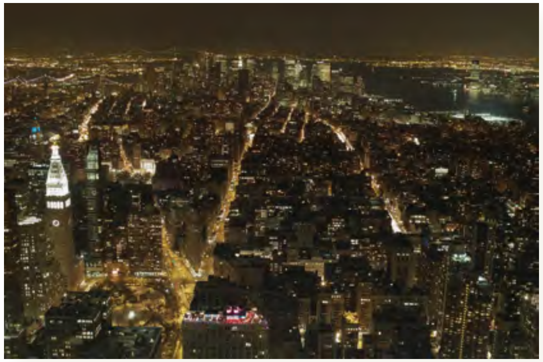
Urban Observatory

Most of the Urban Observatory slides courtesy Greg Dobler

G. Dobler, et al., Dynamics of the urban lightscape. Information Systems, 54:115 – 126, 2015.

the CUSP Urban Observatory

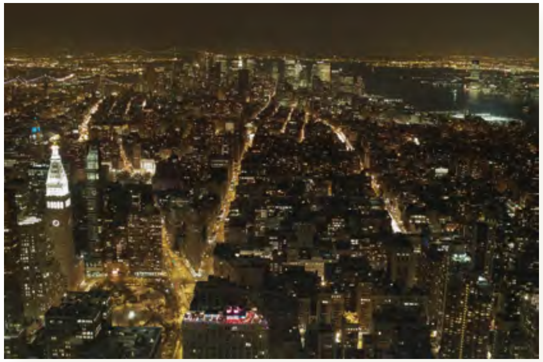
- unique user facility for persistent and synoptic observations of cities with detailed analysis



the CUSP Urban Observatory

- unique user facility for persistent and synoptic observations of cities with detailed analysis

- seek high impact science and applications to enhance public well-being, city operations, and future urban design

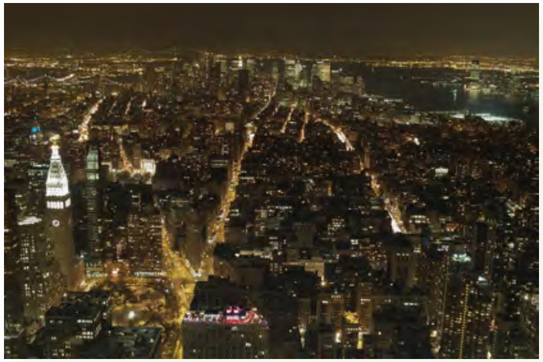


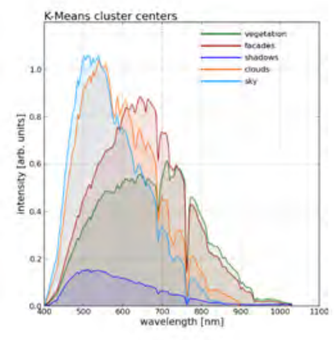
the CUSP Urban Observatory

- unique user facility for persistent and synoptic observations of cities with detailed analysis

- seek high impact science and applications to enhance public well-being, city operations, and future urban design

- instrumentation to include both broad band and hyper-spectral from visible to infrared wavelengths





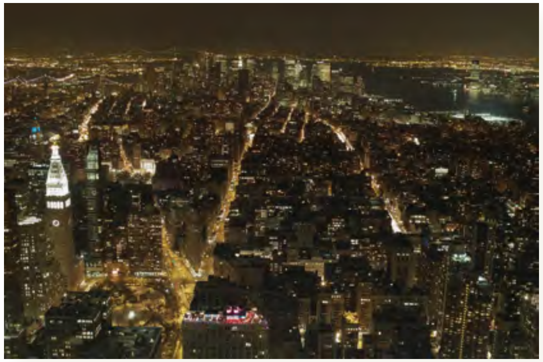
the CUSP Urban Observatory

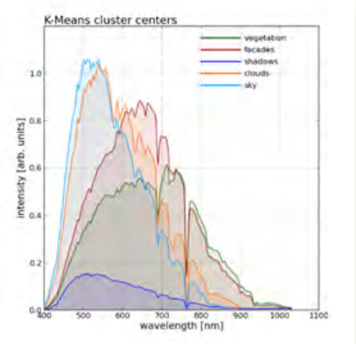
- unique user facility for persistent and synoptic observations of cities with detailed analysis

- seek high impact science and applications to enhance public well-being, city operations, and future urban design

- instrumentation to include both broad band and hyper-spectral from visible to infrared wavelengths

- combine correlative data including administrative records, *in situ* measurements, topography, etc.





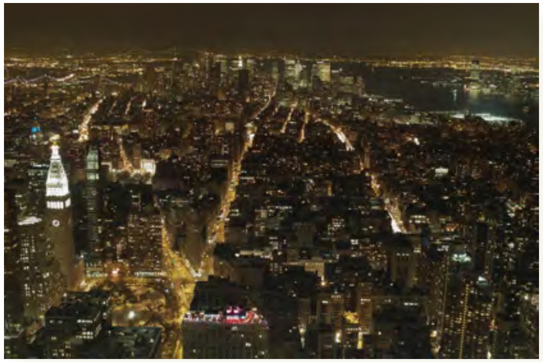
the CUSP Urban Observatory

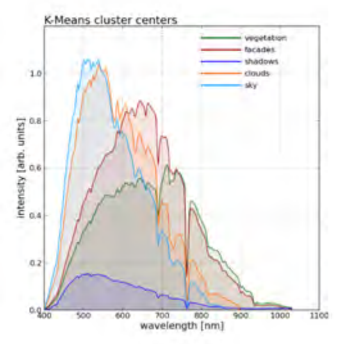
- unique user facility for persistent and synoptic observations of cities with detailed analysis

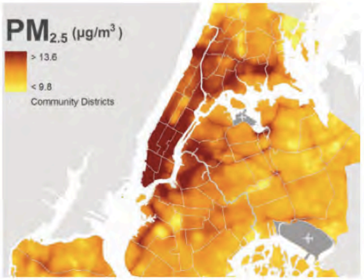
- seek high impact science and applications to enhance public well-being, city operations, and future urban design

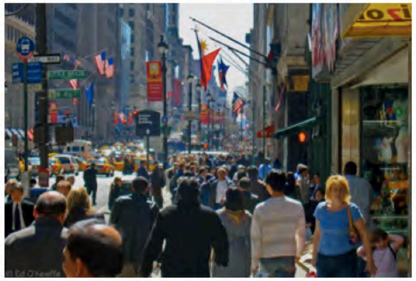
- instrumentation to include both broad band and hyper-spectral from visible to infrared wavelengths

- combine correlative data including administrative records, *in situ* measurements, topography, etc.









Natural Systems
(health of environment)

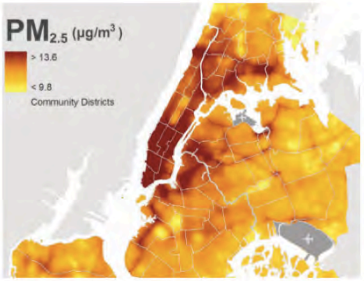
Human Systems
(patterns and behaviors)

Synoptic and *Persistent* Observations of a city: an **Urban Observatory**

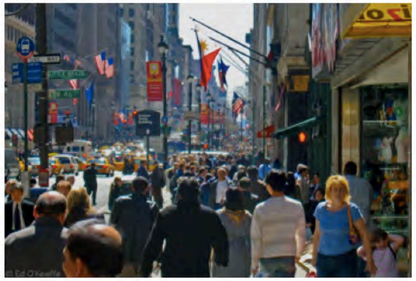
- *Synoptic*: manifesting or characterized by comprehensiveness or breadth of view
- Idea dates to Aristotle:

  "[Aristotle] urged that our view be truly synoptic, a word which had not then become abstract, but was vividly concrete, as its make-up shows : a seeing of the city, and this as a whole; like Athens from its Acropolis, like city and Acropolis together the real Athens from Lycabettos and from Piraeus, from hill-top and from sea. Large views in the abstract, Aristotle knew and thus compressedly said, depend upon large views in the concrete." [P.Geddes, The Evolution of Cities, 1915]

The CUSP Urban Observatory

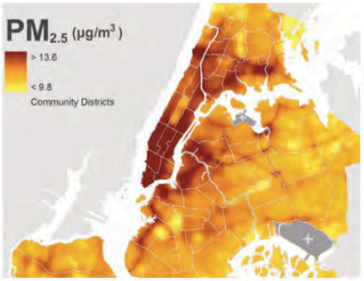


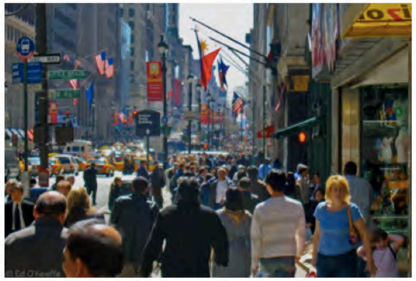
Natural Systems
(health of environment)



Human Systems
(patterns and behaviors)

The CUSP Urban Observatory



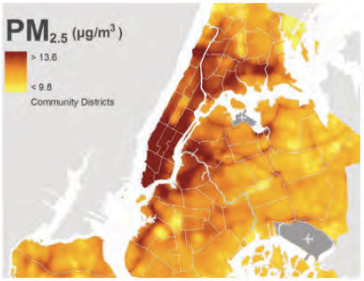


Natural Systems
(health of environment)

Human Systems
(patterns and behaviors)

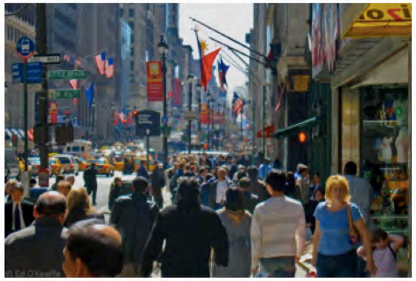
- air quality (hyperspectral, *in situ*)
- fuel type (hyperspectral)
- energy consumption (optical, IR)

The CUSP Urban Observatory



Natural Systems
(health of environment)

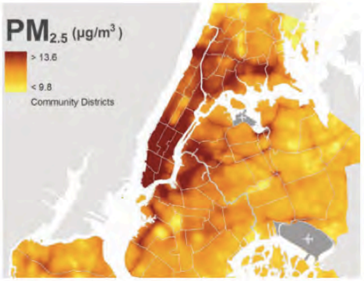
- air quality (hyperspectral, *in situ*)
- fuel type (hyperspectral)
- energy consumption (optical, IR)



Human Systems
(patterns and behaviors)

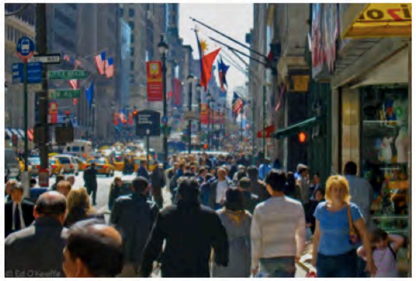
- sleep/wake disturbances (optical)
- emergency response (optical, IR)
- demographic studies (all)

The CUSP Urban Observatory



Natural Systems
(health of environment)

persistent
synoptic
granular

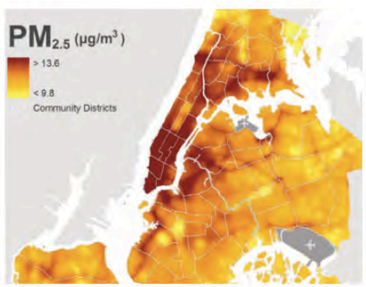


Human Systems
(patterns and behaviors)

- air quality (hyperspectral, *in situ*)
- fuel type (hyperspectral)
- energy consumption (optical, IR)

- sleep/wake disturbances (optical)
- emergency response (optical, IR)
- demographic studies (all)

The CUSP Urban Observatory

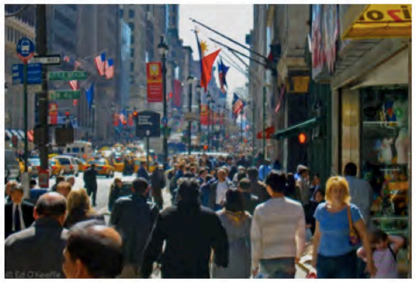


Natural Systems
(health of environment)

persistent

synoptic

granular



Human Systems
(patterns and behaviors)

- air quality (hyperspectral, *in situ*)
- fuel type (hyperspectral)
- energy consumption (optical, IR)

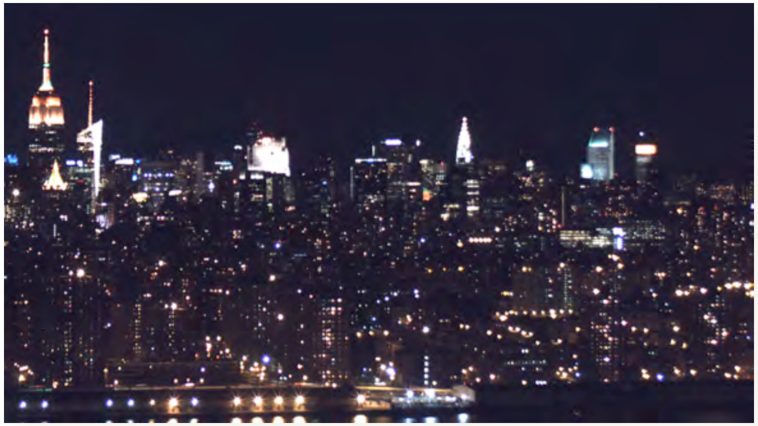
- sleep/wake disturbances (optical)
- emergency response (optical, IR)
- demographic studies (all)

Potential for rapid deployment to other cities

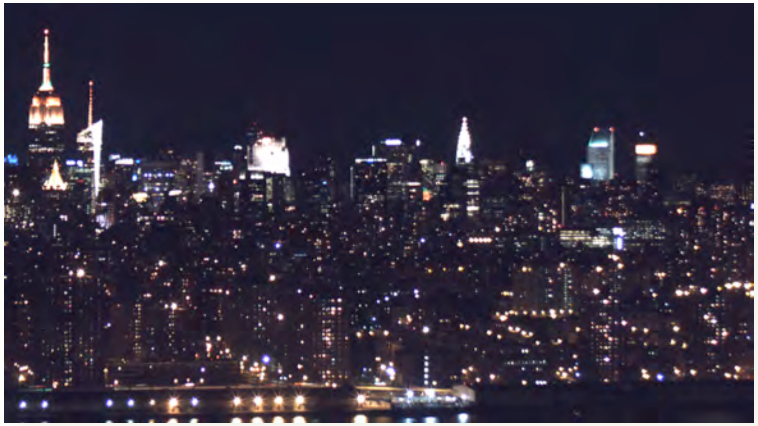
Urban Scenes

CUSP-UO “first light”

Midtown Manhattan: Day & Night



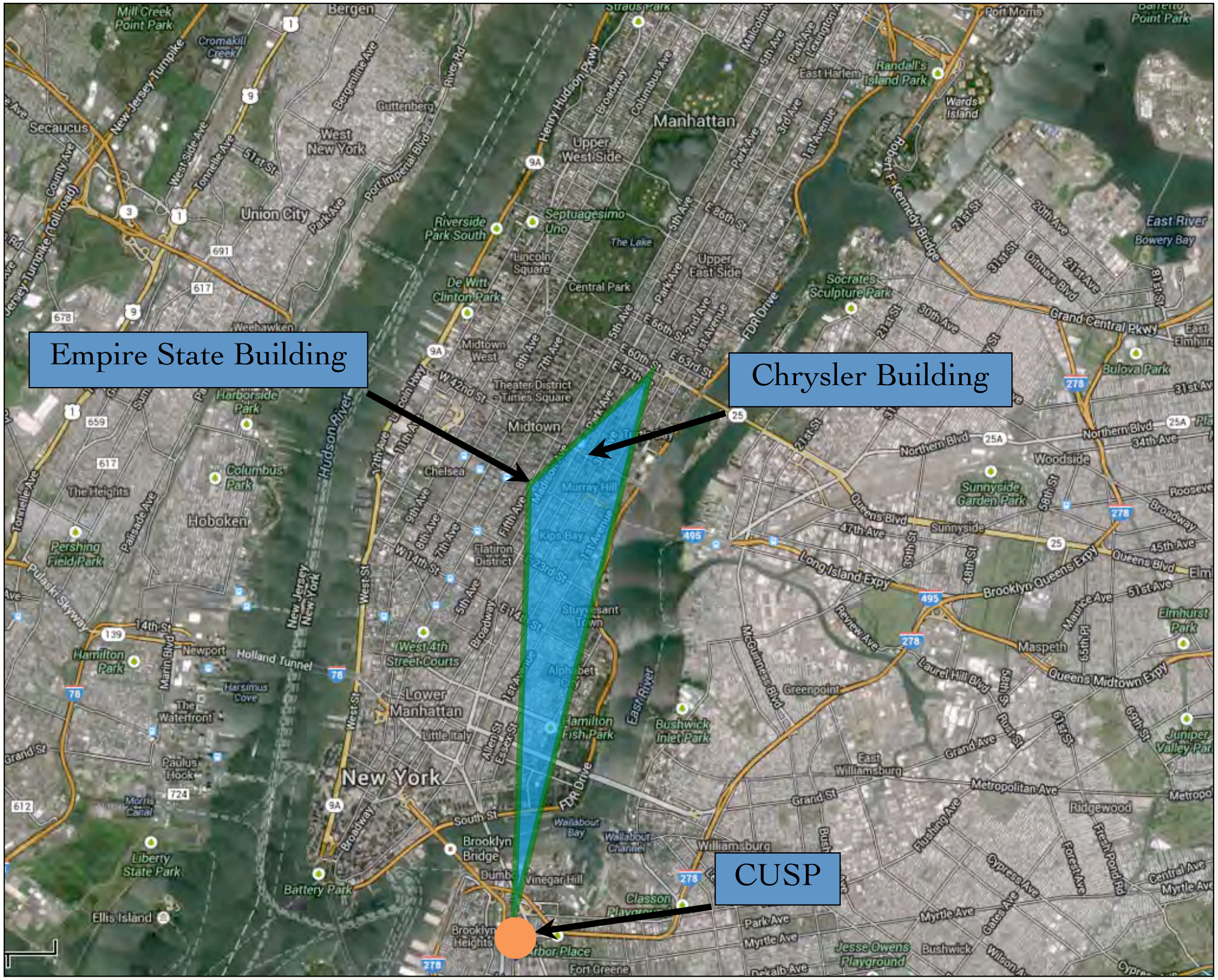
Midtown Manhattan: Day & Night



SHOW THE MOVIE

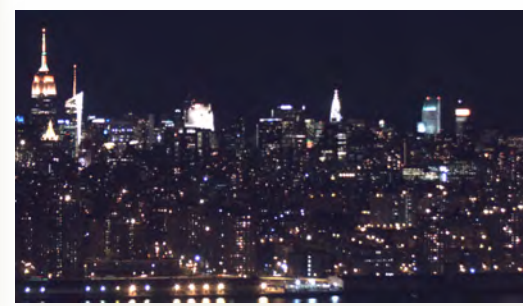
http://www.youtube.com/watch?v=hT3pdUcM75A

Our field of view from 1 MetroTech Center

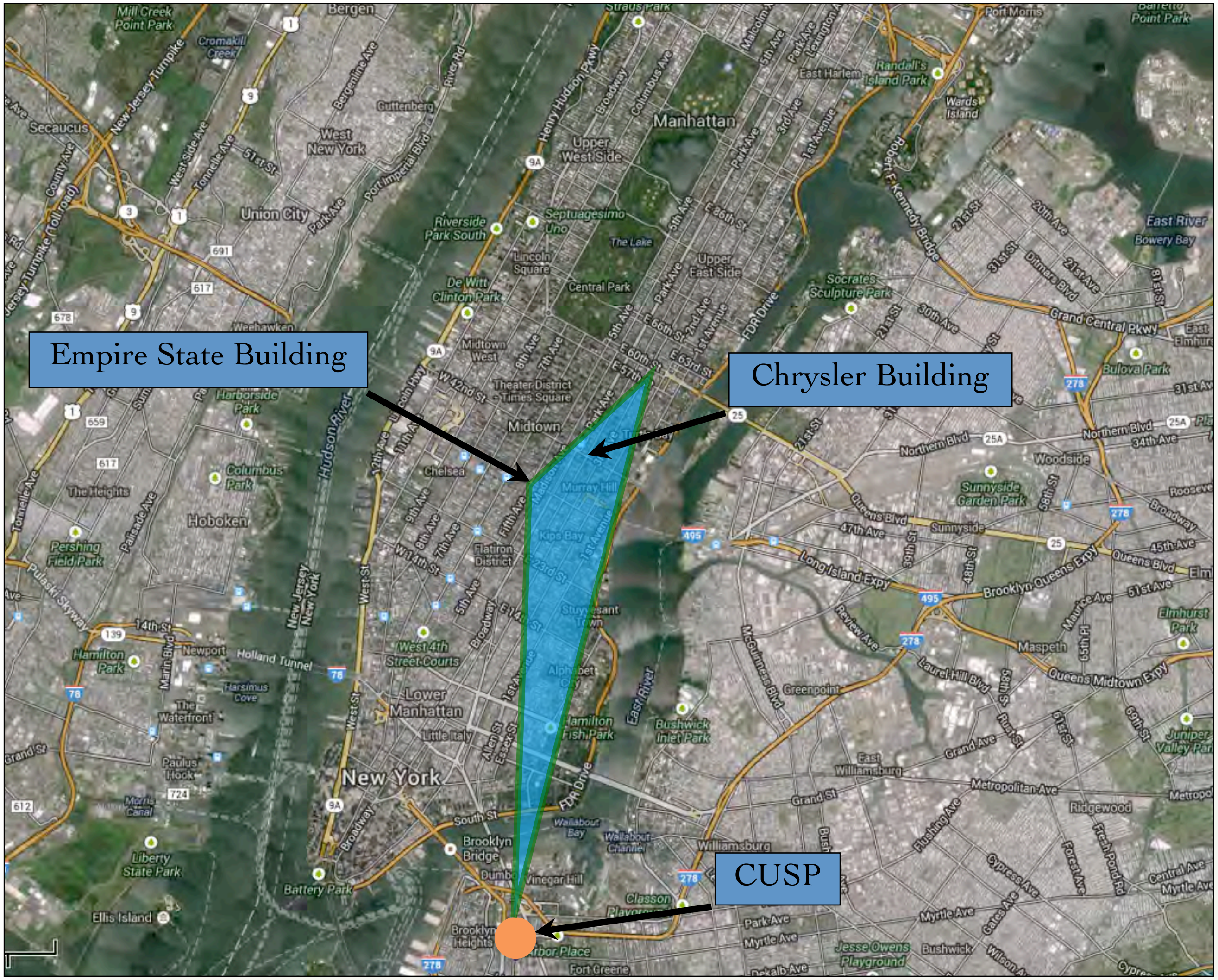


How many city dwellers are in our field of view?

- the population density of Manhattan is roughly 60,000 people per square mile
- our field of view covers roughly 1.7 square miles
- the number of New Yorkers in our field of view is approximately 100,000



Our camera and observations

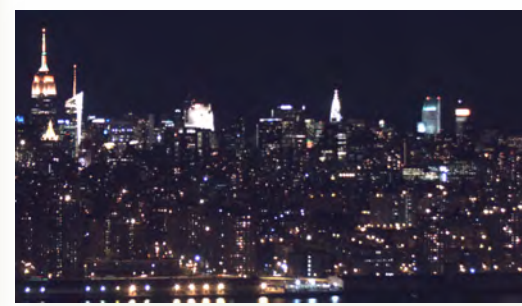


Camera:

- Point Grey Flea 3 USB
- 8 Mega-pixels
- raw image output
- 25mm focal length lens

Observations:

- 1 image every 10 seconds
- from Oct 26 to Nov 16, 2013
- 3 color images at 25MB each
- total data volume ~4.5TB
- custom data processing pipeline



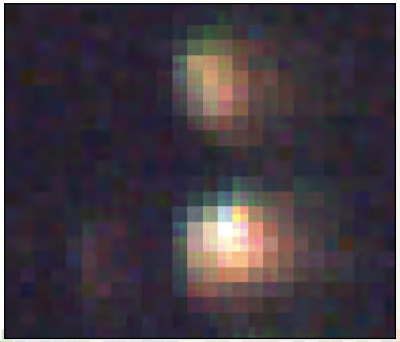
Privacy protections

- Institutional Review Board approval of all projects involving non-open data
- CUSP Chief Data Officer approval

limited # of pixels per window (but atmosphere/instrument effects typically dominate)

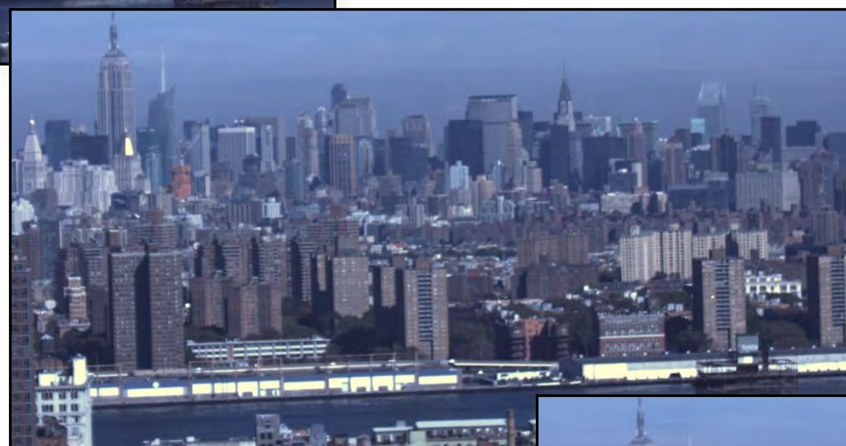
aggregate and de-identified analysis only

Privacy loss through correlated data hard to protect against.



Daytime phenomenology

11:00 AM



11:01 AM

Daytime phenomenology: subtle variations

Daytime phenomenology: subtle variations

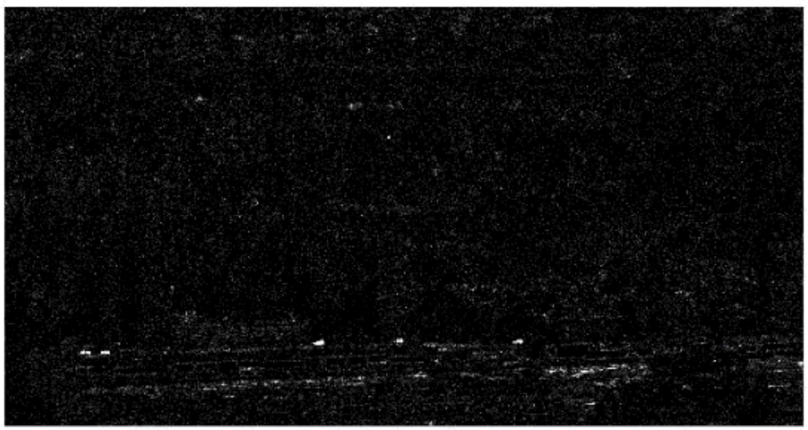
Daytime phenomenology: subtle variations

Daytime phenomenology: subtle variations

animation

Comparison of sequence of images allows for plume identification

raw image



animation

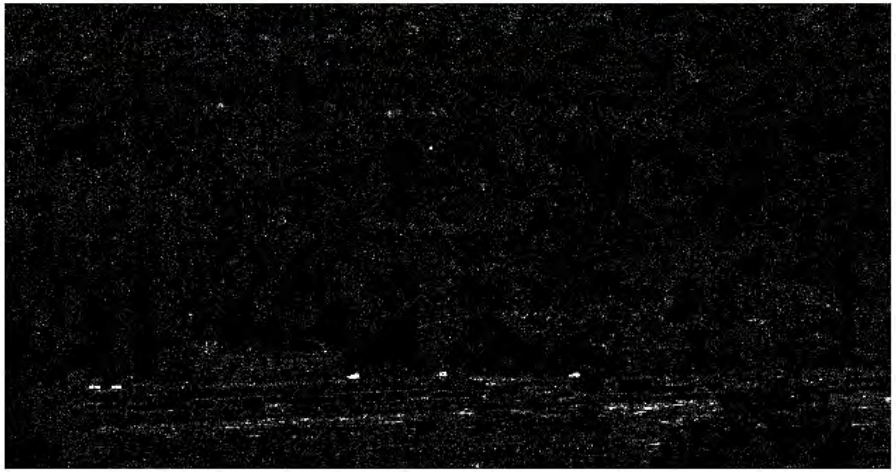
background subtracted

Background subtraction:

- registration to reference image
- form 10 absolute difference images from surrounding frames
- construct the minimum difference image pixel by pixel

Plumes of opportunity

raw image



animation

background subtracted

Background subtraction:

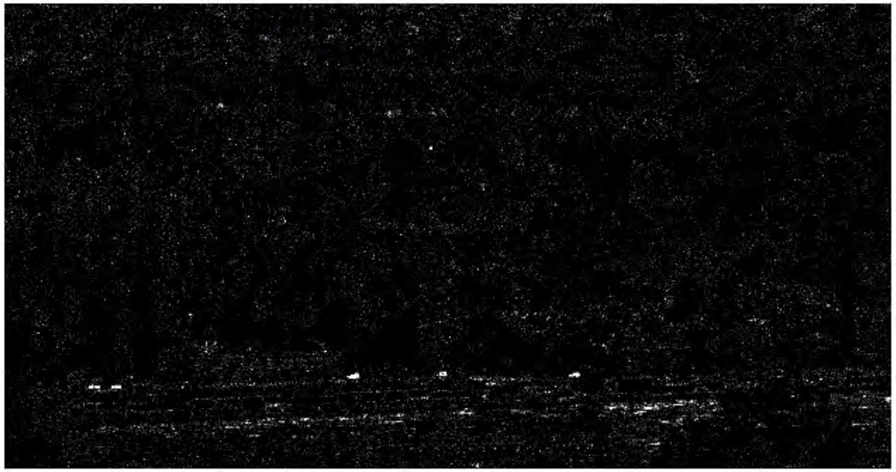
- registration to reference image
- form 10 absolute difference images from surrounding frames
- construct the minimum difference image pixel by pixel

Plume ID & tracking:

- denoise background subtracted image
- identify excess/deficit in luminosity space
- cross check object location in color space
- localization and probability weighted tracking of centroids

Plumes of opportunity

raw image



animation

background subtracted

Background subtraction:

- registration to reference image
- form 10 absolute difference images from surrounding frames
- construct the minimum difference image pixel by pixel

Plume ID & tracking:

- denoise background subtracted image
- identify excess/deficit in luminosity space
- cross check object location in color space
- localization and probability weighted tracking of centroids

Applications:

- **plume rate**
- **repeaters**
- **urban winds**
- **carbon vs steam emissions**
- **TOO (triggered) observations (e.g., hyperspectral analysis of plumes using visible image analysis as trigger)**

More lights are on during weekdays than weekends

Saturday 11:00PM | Monday 11:00PM

lights ON Saturday but not Monday | lights ON Monday but not Saturday

More lights are on on weekdays compared to weekends for both residential and commercial buildings.

Selecting “points of light”

- each frame is “registered” to a common frame by spatial correlations
- 4,200 window apertures are identified by hand (out of approximately 20,000 windows in the scene)

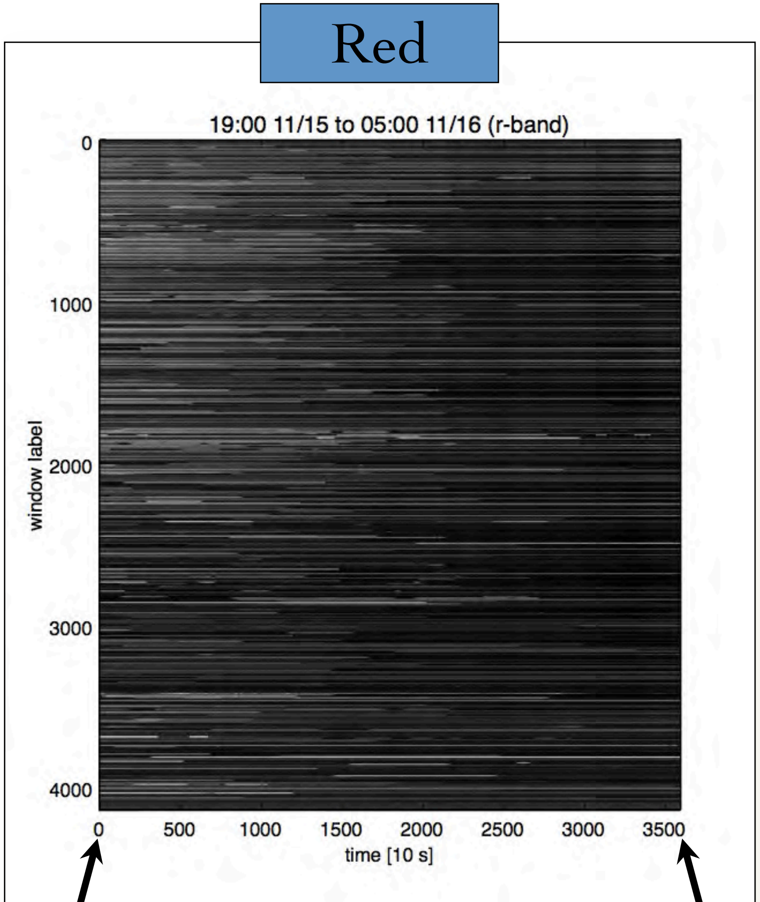
Selecting "points of light"

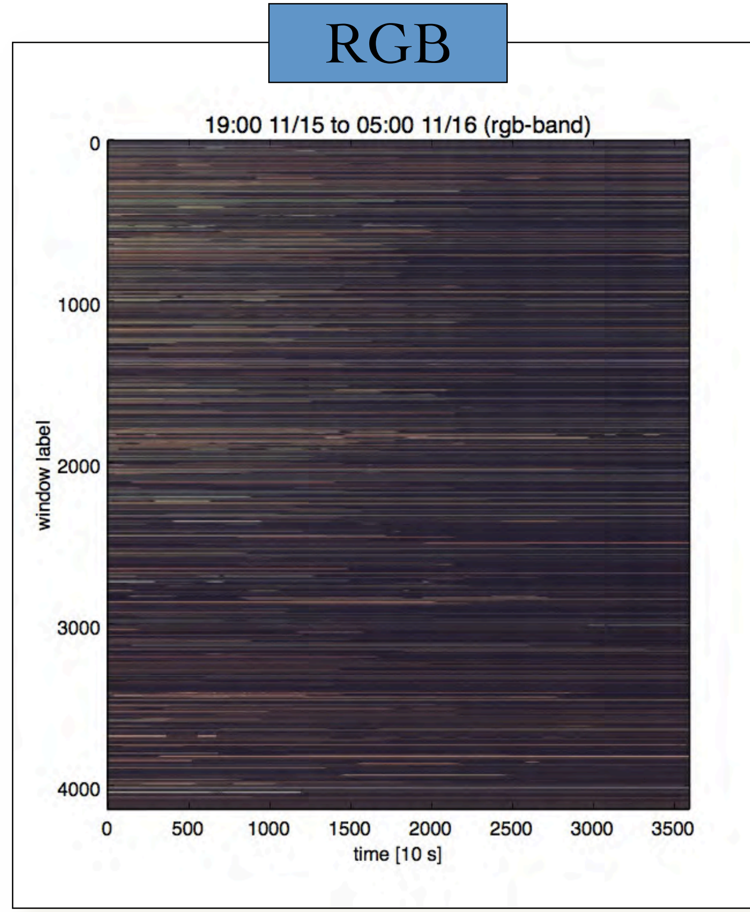
- each frame is "registered" to a common frame by spatial correlations
- 4,200 window apertures are identified by hand (out of approximately 20,000 windows in the scene)

Selecting "points of light"

- each frame is "registered" to a common frame by spatial correlations
- 4,200 window apertures are identified by hand (out of approximately 20,000 windows in the scene)
- for each frame, the average brightness of each window is calculated in 3 bands (RGB)
- the brightness of a given window as a function of time is referred to as its "light curve"

Light curves of windows

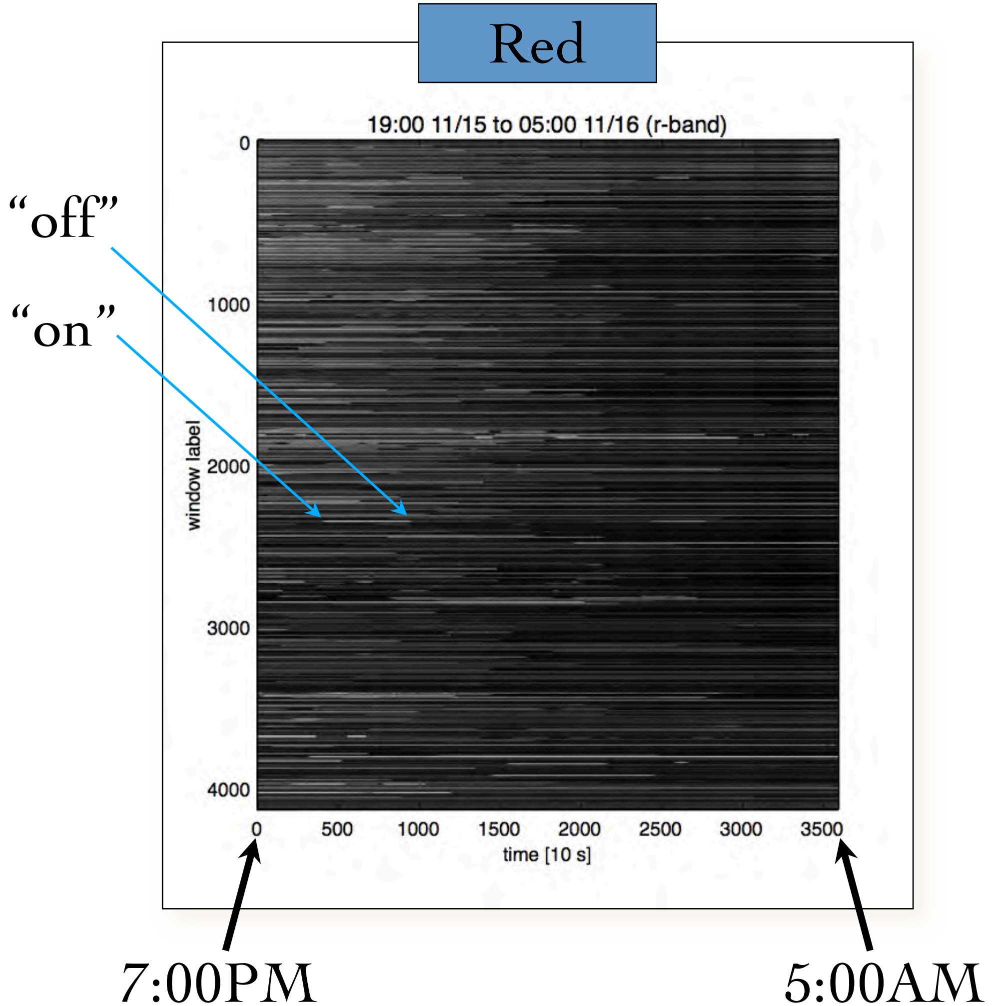


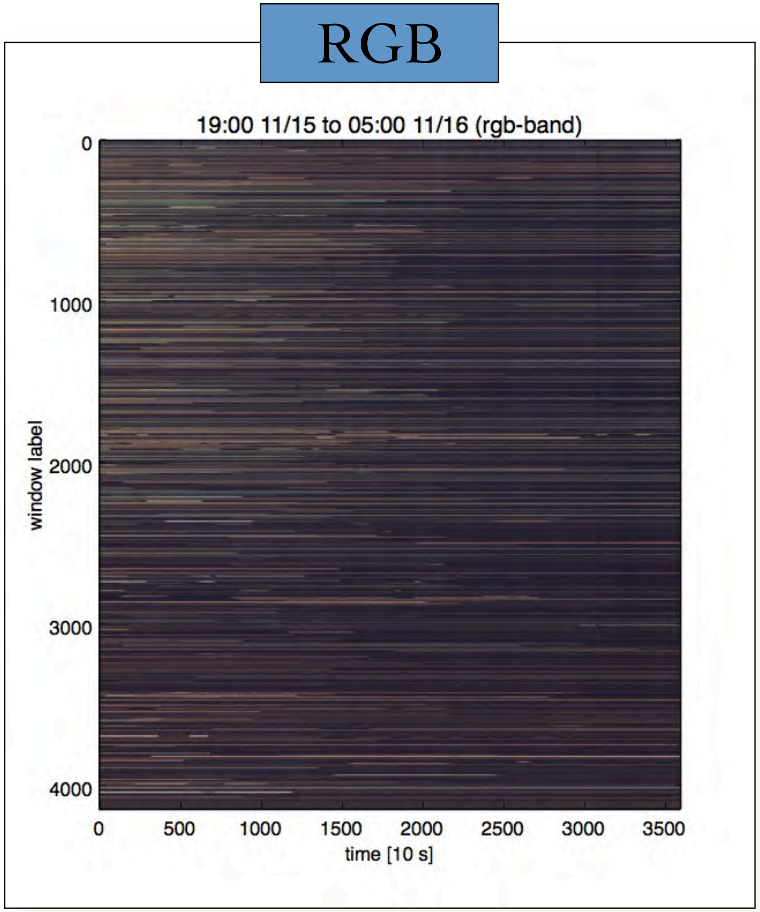


7:00PM

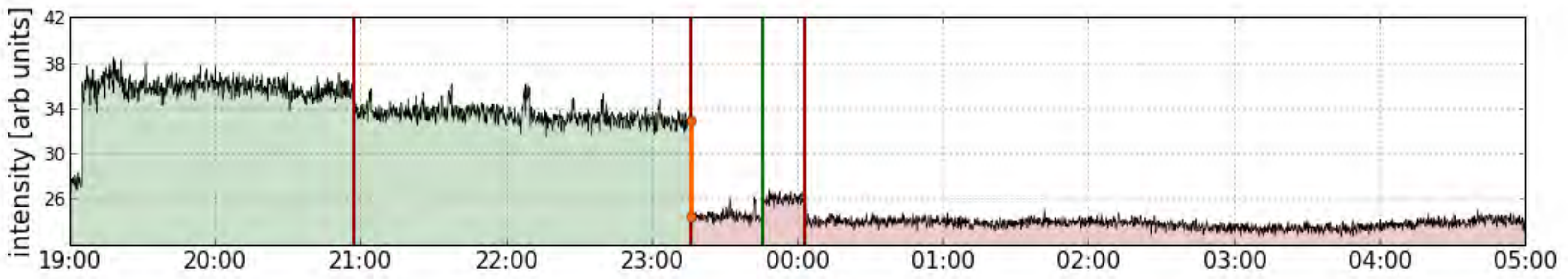
5:00AM

Light curves of windows





Selecting on/off transitions



For a given window light curve, on/off transitions are identified via an automated algorithm:

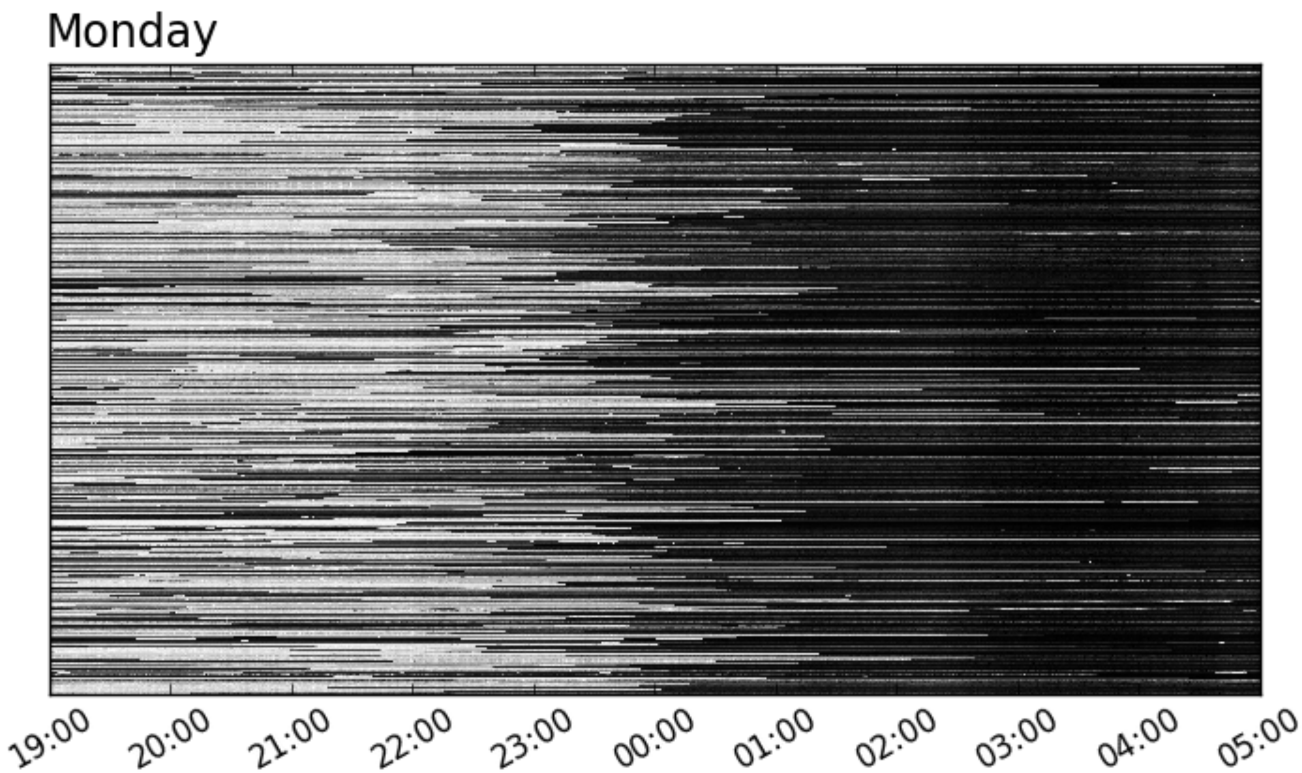
•selects large "jumps" in the light curve relative to background fluctuations

•3-color information and noise statistics are used to improve accuracy

•does not involve any human intervention or interpretation for analysis of individual light curve transitions (i.e., the analysis is de-identified)

In aggregate, on/off transitions are a measure of human (and automated) activity that can yield information about occupancy, behavioral properties of the population, etc.

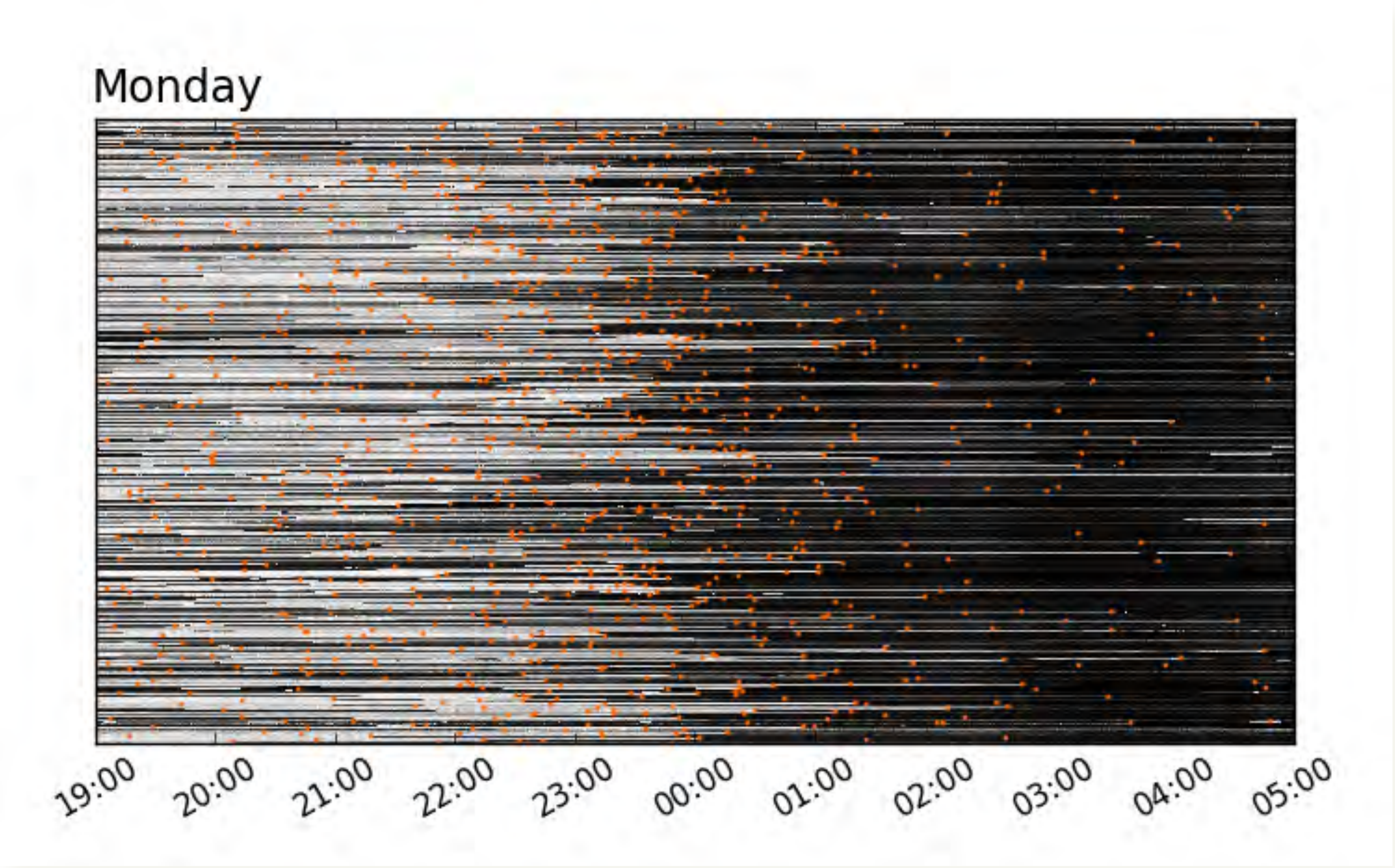
Identifying patterns of behavior



These light curves are reasonably unordered, but is there a discernible pattern hidden in the data?

For example, what if we were to order them according to their “final” off transition?

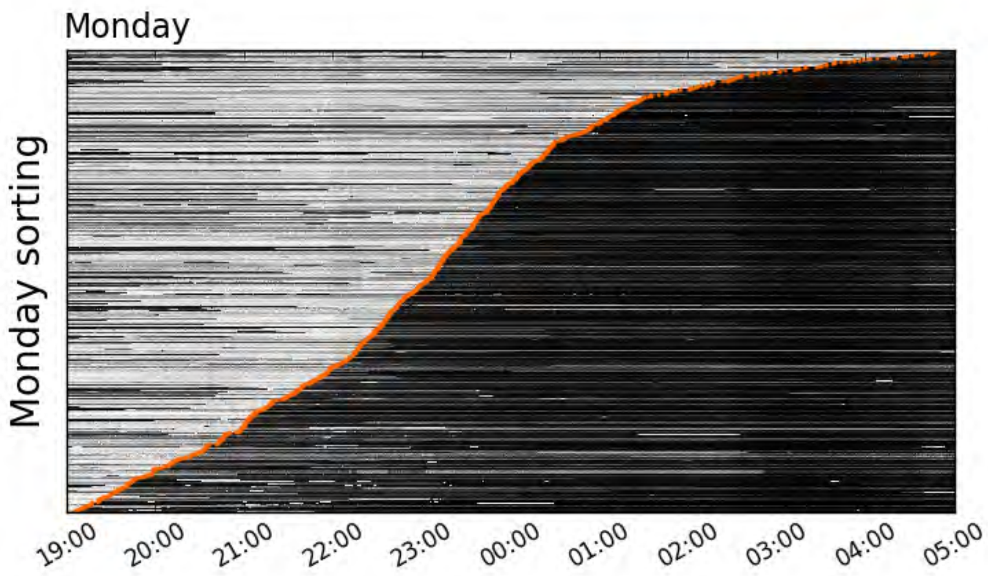
Identifying patterns of behavior



These light curves are reasonably unordered, but is there a discernible pattern hidden in the data?

For example, what if we were to order them according to their "final" off transition?

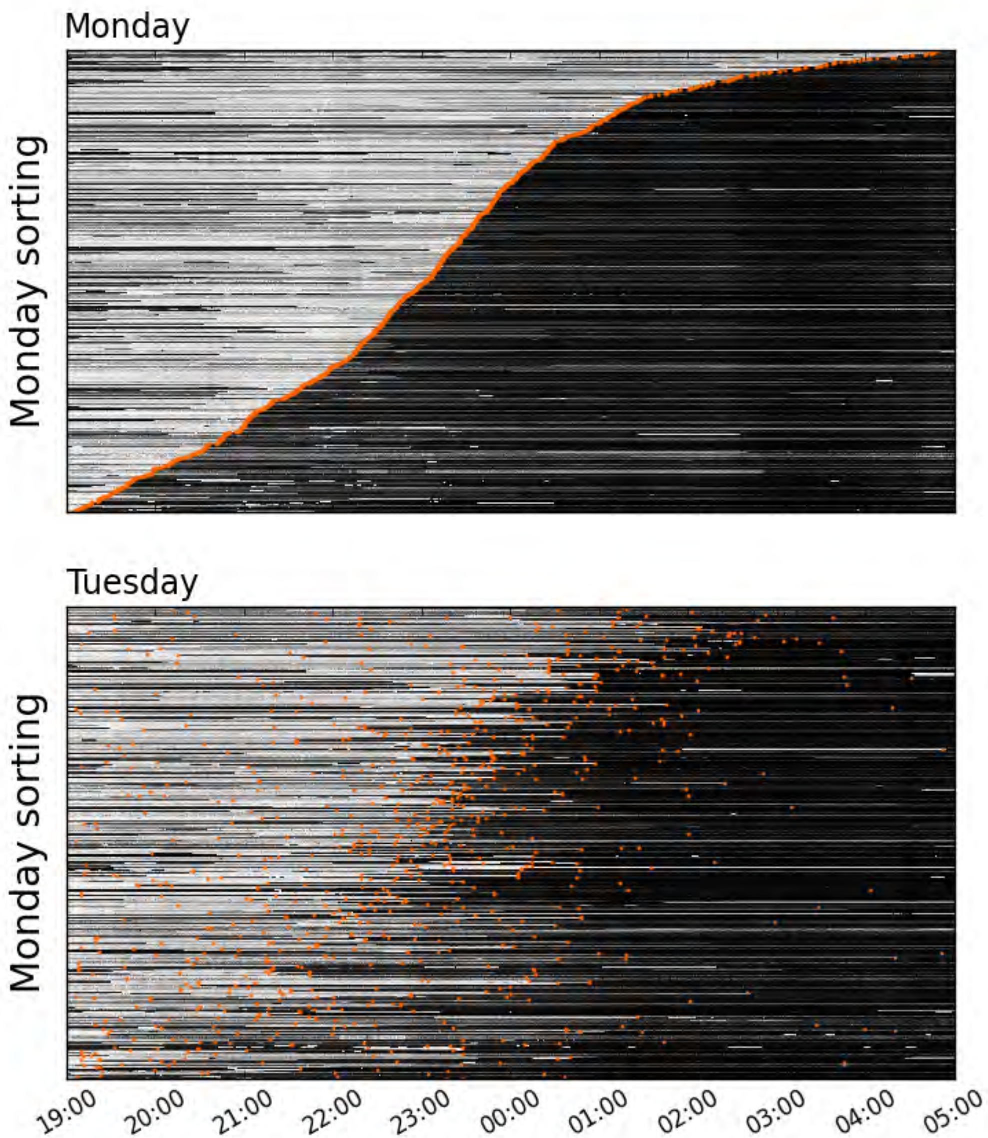
Identifying patterns of behavior



The distribution of off times has a very clear pattern. This is aggregate behavior of the population as measured via observations "from a distance" as opposed to survey (or obtrusive forms of monitoring) methodology.

Does the pattern repeat?

Creatures of habit?

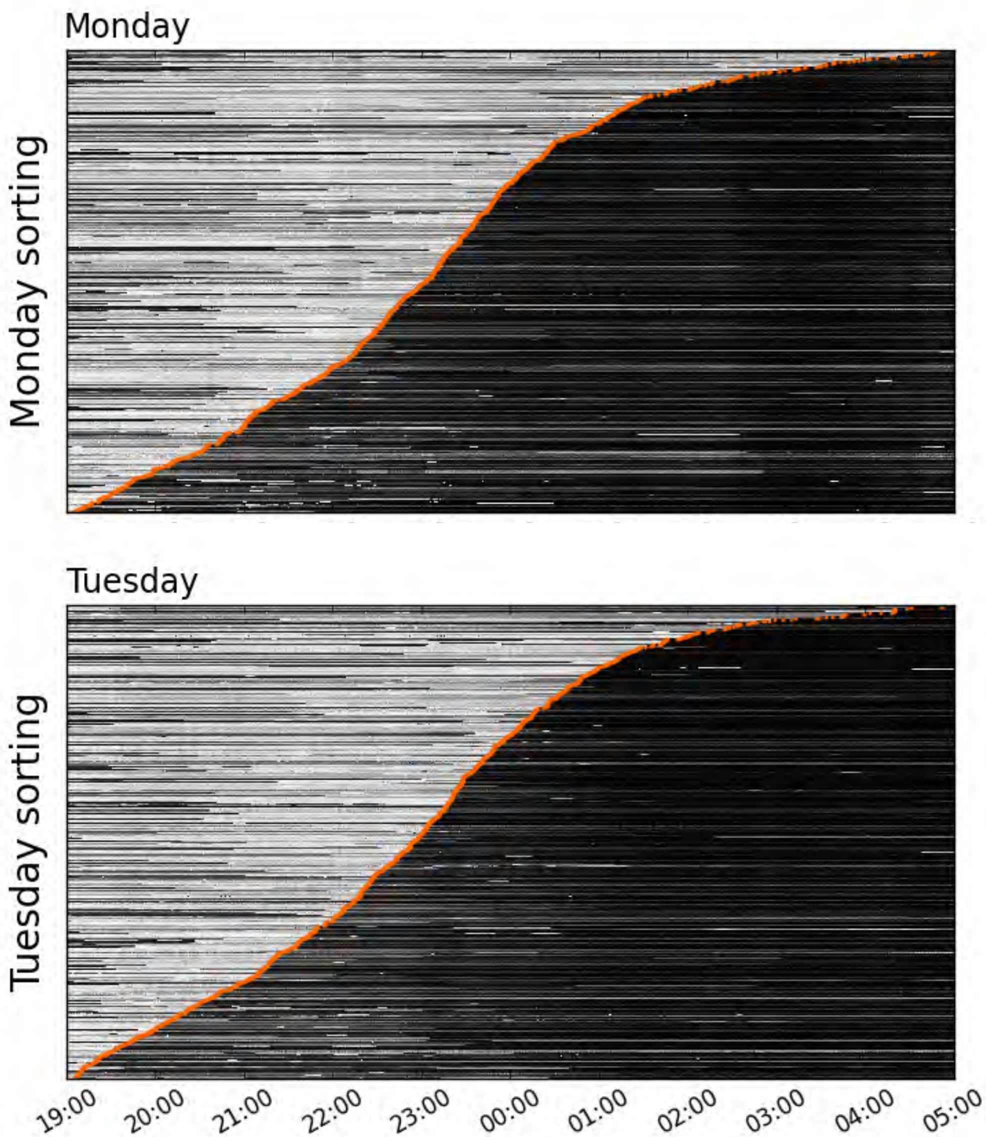


The distribution of off times has a very clear pattern. This is aggregate behavior of the population as measured via observations "from a distance" as opposed to survey (or obtrusive forms of monitoring) methodology. This is data from one observation zone; others views, with other populations in NYC or elsewhere may differ.

Does the pattern repeat?

Answer #1: NO! (weak similarity, though)

Creatures of habit?

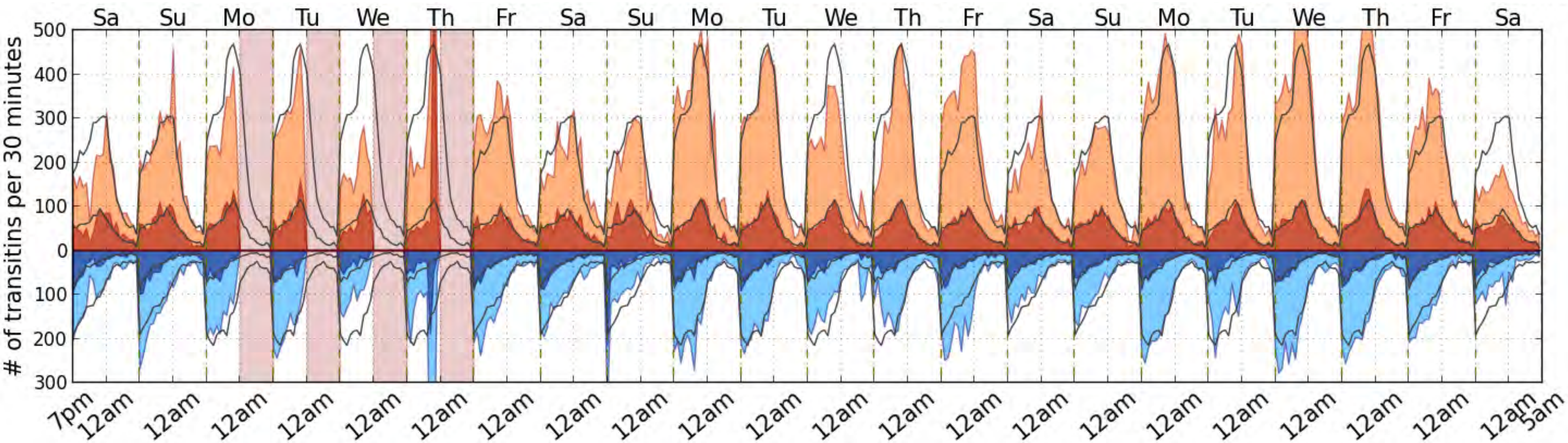


The distribution of off times has a very clear pattern. This is aggregate behavior of the population as measured via observations "from a distance" as opposed to survey (or obtrusive forms of monitoring) methodology. This is data from one observation zone; others views, with other populations in NYC or elsewhere may differ.

Does the pattern repeat?

Not with individual windows, but in aggregate. This is more than law of large numbers.

Pulse of the city lights



The success of this type of urban science inquiry is only possible given the CUSP-UO's unique combination of <u>high coverage</u>, <u>high granularity</u>, and <u>high persistence</u>.

Although we cannot directly infer behavior from on/off transitions alone, the large scale patterns of light are a measure of **aggregate trends** in the activity of New Yorkers (e.g., sleep/wake cycles, (proxy for) energy use, etc.).

The promise and challenge: Use big data wisely. We are only at the beginning. There are great opportunities for innovation and discovery as well as improvements of tradition city operations. Balance needs to be found between using data and protecting privacy.

CUSP Activities include

- Multi-data correlations to improve city resource allocation
- Sound / Temperature / Pollution
- Predictive Policing
- 911/311 call handling
- Fire inspections and correlated building information
- Mobility
- Park usage (people tracking with vis and IR)
- Novel sensing of public health (biogenome of NYC)
- Building efficiency
- Decision science, e.g., pedestrian/car interaction at intersections, quantified community, reaction to events and policies
- Underground economy, e.g., drugs, prostitution, gambling
- Sim City: develop models, study anomalies.
- Infrastructure decisions, allocation of resources